THE CHEAP TRICKS OF COMPATIBILISM AND WHY THE PROBLEM OF FREE WILL WON'T GO AWAY

PAUL DAVIS

DEGREE OF PH.D.
UNIVERSITY OF EDINBURGH
1989

UNIVERSITY OF EDINBURGH

ABSTRACT OF THESIS (Regulation 7.9)

The thesis begins with my arguing that, despite some trends in 20th c. philosophy, there is a perfectly genuine, meaningful (and indeed, extremely difficult) set of problems concerning free will. No amount of clarity or precision regarding our employment of language (I argue) can make problems about free will dissolve into nothing, and in fact, any attempted formalisation of one or other problem about free will is really an attempt at expressing a preverbal worry in language without losing it. In the early part of the thesis I cite what I think are penetrating formalisations of determinism and incompatibilism, provided by Ginet and Van Inwagen.

In Chapter 3, I elucidate and criticise 'language strata' compatibilism, a position most notably held by Donald Davidson, this being the suggestion that the irreducibility of the mental categories secures our freedom. In Chapter 6, I discuss and criticise the empirical compatibilism of Keith Lehrer, and other compabibilist projects which are tried and found wanting include conditionalism (Chapter 8), and the Mind argument (Chapter 10), though I recognise that the latter is really an argument for the incompatibility of free will and indeterminism.

The empirical attempt at establishing the absence of free will is also criticised (Chapter 7), as is Frankfurt's attempt at marginalising free will/determinism as anything affecting "freedom" (Chapter 9), as well as the eliminative materialist's effort at transforming all our talk about the mental (Chapter 5).

Having argued for incompatibilism, the concluding chapters of the thesis include my arguing that determinism looks unlikely and that, whilst very serious difficulties remain, any positive theory of free action will turn on the possibility of causation without determination and this, in turn probably, on the possibility of 'agent-causation'.

My final chapter is a brief criticism of Frankfurt's argument that moral responsibility does not require free will.

CONTENTS

1. THE ESSENTIAL PROBLEM

In a footnote to his essay "One Determinism"[1], Ted Honderich expresses some curiosity regarding both the number and the eminence of contributors who have pleaded ignorance of the content of the thesis of determinism. J.L. Austin, for instance, was 'inclined to think that determinism...is a name for nothing clear' and that it 'has been argued for only incoherently' [2] and , in an equally celebrated work, P.F. Strawson aligns himself with 'the party of those who do not know what the thesis of determinism is', although, as he says, he has 'some inkling - some notion what sort of thing is being talked about[3]'. Honderich then goes on to make the deeply revealing point, I think, that both of the aforementioned distinguished essays arrive at substantial conclusions about what was thought to be unclear. I would myself feel perplexed about this quirk in the claims and conclusions of the likes of Austin and Strawson, were it not for what I think it does in fact reveal. It is not that I believe that the thesis of determinism cannot be stated formally, or that it cannot be stated without incoherence, or even (as will all be made clear) that the worries traditionally supposed to be grounded in the prospect of its truth, are bogus. In fact, I will be arguing that each of those claims ~~are~~ is mistaken. However, what is, I think, lying at the bottom of the prima facie incompatibility between this claim and conclusion of Austin and Strawson, the claim of ignorance about what determinism is, and the substantial conclusion in the face of this ignorance about determinism, is actually something fundamental about philosophical problems in general. Indeed, this feature of philosophical problems is so pervasive, and so essential to what it is that make philosophical enquiry at all valuable, that I am surprised at the likes of Strawson failing to recognise it. The point is this: despite what I think is the not only mistaken but intellectually repressive idealist approach within much recent philosophy, which dismisses a great many philosophical problems as mere manifestations of our particular historical situation or of the accidental forms of our language, the sources of philosophy are preverbal and often precultural and, as Nagel points out in his introduction to

The View from Nowhere, one of its most difficult tasks is to express unformed but intuitively felt problems in language without losing them. It is therefore not only entirely unwarranted of the likes of Strawson and Austin to conclude from the fact that they do not have a clear-cut, straightforward statement of the thesis of determinism at their fingertips, that such a formalisation is impossible without incoherence and that there is no free will/determinism problem, but indeed, this is entirely of a piece with most philosophy. It can therefore be no surprise, and in fact again I regard it as of the utmost significance, that Strawson has 'some inkling' of what kind of thing determinism is. To use Nagel's language, Strawson "feels" the free will/determinism problem, I'm sure, every bit as much as those who don't take his 'much ado about nothing' attitude (otherwise it's doubtful whether he would have ever bothered writing "Freedom and Resentment"). And, as I've been trying to point out, we often do start out only with a 'felt' inkling of a philosophical problem, and then we must struggle with existing concepts to formalise it in a way which we can recognise as a satisfactory articulation of this preverbal, or at least inchoate worry. To take another example, consider realism/idealism. This seems to me to quite obviously have its source in a highly primitive, 'gut' kind of worry, which we must hope that we can get our concepts to meet. No degree of Wittgensteinian purity can make it go away, since it is an issue which outruns any level or dexterity of linguistic therapy. We can, for instance, spend any amount of time we wish or achieve any level of precision we could, on (say) criteria of application of material-object words, under what conditions we are justified in making the jump from sense-experience to perceptual claim etc., and yet the questions which comprise realism/idealism would not have gone away, i.e. are any of the qualities we ascribe to objects essentially subject-dependent (this position being famously espoused, of course, about the 'secondary' qualities)? If so, which ones? Could they all be? Can we make sense of the notion that the entire world could be completely different from how it appears to us and if so, is there any reason for believing that this is the case? Could the ultimate constituents of the universe be beyond the grasp

of any possible human conception? Can we make sense of this idea? (Nagel, of course, believes that, not only is the idea coherent, but that the universe probably does outrun any possible human conception of it.) Quite regardless of the contingencies of language or culture at any time or place, these questions remain, and so there is a realism/idealism problem which we must then do our best to articulate. Similarly with ethical or aesthetic objectivity: however difficult it may seem to articulate the problem, it would take a very great deal indeed to persuade us that there is not, at the very least, a different kind of subject-dependence and subject-independence in ethics and aesthetics than there is in, say, physics, (despite Bambrough's efforts in Moral Scepticism and Moral Knowledge). And so on. As Nagel also points out, certain forms of perplexity (amongst which he includes freedom) seem to embody more insight than any of the supposed solutions, and, in his essay 'Responsibility for Self'[4], Taylor makes the point most succinctly:

> In philosophy typically we start off with a question, which we all know to be badly formed at the outset. We hope that in struggling with it, we shall find that its terms are transformed, so that in the end we will answer a question which we couldn't properly conceive at the beginning. We are striving for conceptual innovation which will allow us to illuminate some matter...which would otherwise remain dark and confused.

Indeed only just prior to the start of the above extract, Taylor suggests that anyone who has struggled with a philosophical problem is aware of this kind of feature, and it is surely reasonable to assume that this includes Strawson. All the more reason, I think, to regard it as ill-conceived on the part of Strawson (as I will argue in more detail later) to think (as he seems to) that by hinting, through his own proclamation of ignorance, that there is no genuine content to the thesis of determinism and by offering what I take to be some banalities and trivialities, the free will problem will be seen to be bogus.

As the remarks of Taylor cited above suggest, formalisation of a

philosophical problem frequently doesn't come easily. And indeed, before we can even arrive at what we know to be an unsatisfactory articulation, we may well have to first articulate and recount what gives rise to the feeling that there is a problem at all. I have already sketched an example of this, in the case of realism/idealism, citing the fundamental features about ourselves and the world which gives rise to the problem. Similarly with free will/determinism. We can only arrive at even a poor and murky articulation of the problem through articulating and recounting factors which give us the primitive feeling that there is anything to be worried about at all. It strikes me, indeed, as wholly unsurprising that in 'Freedom and Resentment' Strawson makes no serious effort to consider these relevant factors, and it is little more than trite to suggest that had he done so, he would never have thought that the issues which he claims to address could have been dealt with in the dismissive fashion he displays.

In common, again, with a great many philosophical problems, free will/determinism has something of a chequered history. In the same way in which it may come more easily to say what it is (for instance) that empiricism denies rather than what it asserts, we could fairly painlessly grant the likes of Strawson that the terms "free will" and "determinism" do not each, within philosophical history, mark off single unambiguous theories going up for grabs. To see this, however, as a reason for denying the authenticity of the free will/determinism problem would also commit one to denying the credibility, I expect, of most classical problems of philosophy, e.g. empiricism/rationalism, realism/idealism, monism/dualism, and perhaps most famously of all, the analytic/synthetic distinction. I expect that neither Strawson nor Austin would want to do this. Here again, free will/determinism strikes me as merely typically philosophical; the various attempts throughout philosophical history at stating both theories have been attempts, for better or worse, at stating some fundamental intuition, itself felt in the most deep and primitive way, and resulting from observations, intuitions, reflections and worries about ourselves, the world, and the interaction between ourselves and the world. It can hardly be

any surprise that the free will problem has the (typically philosophical) history which it does have and indeed, the situation is perhaps exacerbated in the case of free will/determinism by the historical presence of the theses of compatibilism and incompatibilism; the former thesis is that human freedom can be reconciled with determinism, the latter that determinism precludes human freedom. Again unsurprisingly, it is frequently the case that not much is specified regarding determinism, and nor is anything specified regarding free will. And in fact, the terms "free will" and "determinism" have been frequently defined so that one explicitly excludes the other, this taking incompatibilism for granted. However, I say again that, at the bottom of all the ceremony and flailing about, are primitive intuitions, and before I cite or consider any specific formalisations of the determinist theory, it is more than worthwhile, I think, to recount what it is that gives rise to those intuitions in the case of free will/determinism. At any rate, to do so can do no harm, and I would hope that it would also make it clear why I regard any of the formalisations of determinism, indeterminism, libertarianism, compatibilism or incompatibilism which I will cite, as challenging and interesting. However, before actually recounting the significant factors, I think it very worthwhile to note comments made by Peter Van Inwagen, in his excellent Essay on Free Will, comments which are at least germane to this point. I began by noting Honderich's slight exasperation at claims made by such as Strawson and Austin which (if sound) would seem to marginalise determinism as any sort of issue, and Van Inwagen displays a similar kind of exasperation about a feature of incompatibilism's fate in philosophical history. It is, notes Van Inwagen, surprisingly hard to find any arguments for incompatibilism, i.e. surprising in light of the fact that arguments are so easy to give:

> Perhaps the explanation is simply that the arguments are so obvious that no one has thought them worth stating. If that is so, let us not be afraid of being obvious.

Van Inwagen then goes on immediately to state what he takes to be an

obvious argument (in that it should occur) for incompatibilism (the Consequence Argument), which he fleshes out admirably in the book, and about which I will go on to say much more:

> If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us.

Van Inwagen's rather brazen remarks here are interesting, I think. Like the efforts of Honderich, they are a sharp reminder to the likes of Strawson and Austin that there may well be worries regarding a thing called 'determinism' and a thing called 'free will'. At the same time, however, they do not gainsay my earlier remarks regarding the primitive, preconceptual well-spring of the problem. Strawson and Austin may say (as I expect a great many others would) that the above formalisation is ludicrously bland, and on inspection can be seen to not actually state any problem, since it contains concepts which are ultimately vaccuous, e.g. 'law of nature'. Even Van Inwagen, however, is perfectly well aware that a great deal of cashing out and defence of this argument is required, a great deal of the book being devoted to just this (including remarks on the notion of a 'law of nature', a notion which I will say something about later). Van Inwagen knows that this formalisation of the Consequence Argument can only be a start. It is the first primitive efforts and, although these embryonic steps may be obvious the subsequent steps (as the book bears out) may be somewhat more tricky. The point is, however, that despite the dismissiveness of Strawson or Austin, it may well be the start, the first efforts, the first tottering steps of something, and if it is not, then we can really only find out by examination and argument (Honderich is also, likewise, well aware that a defence of his determinist premises is required, and so too is Ginet, as I will document).

At the forefront of the factors generating the free will/determinism worry is undoubtedly the achievements of the sciences, and most importantly of the aspiring sciences of man. Strawson may or may not be right in believing that we cannot but assume in most everyday life that on many occasions we are free agents, able to do or abstain from doing this or that at will. This assumption, or alleged assumption of ours could conceivably be a merely local feature, peculiar to a particular historical standpoint, or a particular culture - specific psychological trait. However, there's little to be lost, I think, in allowing for the moment that such an assumption is as given, as fixed as the likes of Kant and Strawson claim. The first problem, however, for this assumption as stated here, is not, in fact, one which is specifically generated by the achievements of science, though it may well be the case that the latter contribute further to its appearance of intractability. This is the question of precisely what an 'ability' or a 'capacity' is: it is said that we believe ourselves able to do things which we don't do. But what exactly does this mean? What do I mean when I say that I can do something at a time when I am not doing it, and how do I know that it is true? As Dennett puts it in a footnote to his essay "On Giving Libertarians What They Say They Want "[5]

> How can the unique four-dimensional world-worm that comprises all that has happened and will happen admit of a notion of possibilities that are not actualities? What does an opportunity look like when the world is viewed sub specie aeternitatis?

I will be arguing later, for instance, that Keith Lehrer, in 'An Empirical Disproof of Determinism?'[6], and M.R. Ayers in 'The Refutation of Determinism , are operating with a significantly evasive notion of capacity, which they don't appear to recognise as a problem. At any rate it appears to me to be a very real problem. And, as I've indicated, the achievements of the sciences tend, if anything, to exacerbate this problem; even if we are confident that we can make sense of the notion of an opportunity, the notion that

we can do something at a time when we are not doing it, there remains the further question of whether we ever actually have opportunities, i.e. whether we actually can genuinely ever do something other than what we do. It may seem to be both a presupposition and implication of the achievements of the sciences that there are, in truth, no genuine alternatives; and that everything, human conduct not excluded, really happens necessarily, and of course, it is well known that Kant was aware of this, and thought himself able to secure human freedom only through postulating a mysterious and inaccessible component of the self, i.e. the noumenal self. The problem is that mechanisms are regarded as the paradigm case of the cause-effect relation, the archetypal exemplars of notions of natural law and inevitability. Once wound up properly and whilst in good working order the clock cannot but go round - it is nonsense to entertain the prospect of its doing otherwise. Furthermore, it would be an especially bizarre piece of nonsense to ascribe the notion of choice to the clock, to say that it is "deciding" to go round for another few hours or, for that matter, to ascribe feelings to it, say of enjoying going round. This sort of consideration both reinforces and is reinforced by the legacy of Cartesianism, according to which mechanisms don't choose, will, think, feel or do anything "mental" (except perhaps in the flimsy metaphorical sense), with the consequence that, in so far as humans do these "mental" things, they cannot be mechanisms. They may, by all means be mechanisms, but they cannot be mechanisms and nothing else besides. This is the force behing the locution "mere mechanism", and the attendant worry, which arose as science seemed to be bringing increasing phenomena under its mechanistic tutelage, the anxiety being fed that it was only a matter of time until humans were, just like the rest, completely conquered; at which point, being no different really from the clock, we would have to abandon grand ethereal notions that we are free to choose what to do and what not to do. As Dennett puts it:

> How can a material thing (a mechanism) be correctly said to reason, to have reasons, to act on reasons?[7]

At any rate the relatively skeletal set of considerations I have just advanced seem to me, despite Strawson, to be more than enough to generate scepticism regarding human freedom, and the ascriptions of responbility which seem to depend on it, this in turn seeming to be a condition of a whole set of reactive attitudes, such as praise, blame, love, gratitude, indignation, and, of course, resentment. These considerations should give us an idea of what a rigorous formalisation of the thesis of determinism would look like and again, it has to be significant that Strawson doesn't look at one actual formalisation. Three particular formalisations of the determinism thesis have struck me as admirably concise and challenging, with nothing obviously bogus about the consequences of their truth for the way in which we view human actions, these being Carl Ginet's "Might We Have No Choice?"[8], Ted Honderich's "One Determinism"[9], and that put forward by Van Inwagen (the aforementioned) in his Essay on Free Will. In fairness to Strawson and Austin, I suppose the title of Honderich's essay contains a tacit recognition that the history of writing on the free will problem displays its share of conceptual confusion and possibly even downright subterfuge (though again, I'm not convinced that this is appreciably more so than with any other traditional philosophical problem), hardly forgetting the fact that the bare facts of timing leave it unreasonable to hold the Strawson of "Freedom and Resentment" or the Austin of "Ifs and Cans", responsible for failure to be acquainted with the cited works of Honderich, Ginet and Van Inwagen. However, the fact that such formalisations are possible at all reveals, I think, that it is unwarranted of Austin or Strawson to claim that determinism is some sort of non-thesis or non-problem, which a proper appreciation of our practices and conditions of language use would reveal to be empty. I will soon devote some attention and consideration to Ginet's enterprise, in particular. The root problem to be formalised is surely that of the threat to personal authorship: the prospect of an all-encompassing explanation of our actions in terms of cause and effect, à la the physical sciences, forces upon us the primitive, but deeply felt intuition that we are, in fact, merely the focus of causal summation for external influence, there being no room left for me as the

author of my decisions and actions. If a person is one physical object among many in the natural order of cause and effect and can (in principle at least) be exhaustively described in such terms, then how can the notion of a person as decision maker get a foothold? That is, if a person is absolutely nothing more than a cog in the cause-effect wheel, then surely what the person will do can be wholly accounted for in terms of natural causal law, thus rendering the person ultimately impotent with regard to what he will do. It will be no surprise that this has by no means been accepted by everyone, i.e. not everyone has accepted that the prospect of a comprehensive physical (causal) explanation of all human activity has the desperate consequences for our notions of freedom and responsibility that it is generally taken to have. Among those dissenters, as I will elucidate, are Davidson, Dennett and Honderich. However, as I will also be elucidating in what is to come, I agree (as I will document shortly) with the claim endorsed by Malcolm[10], Van Inwagen and Ginet, that comprehensive physical explanation of all human activity would cut across all ontologies and would leave no room whatever for notions of freedom and responsiblity; this, despite, not only Kant, but the desperate manoeuverings of 'language strata' theorists such as Davidson, 'human nature' apologists such as Strawson, and those who try to purge causality of its full-blooded centrality and significance, such as Dennett and Harry G. Frankfurt (as I will cite later). And, of course, hardly to be forgotten is the notion of the self. If it is taken to be a different type of object from, and ultimately independent of, the natural world of cause and effect, it would seem impossible to fit it in anywhere, if a comprehensive physical explanation of all human activity can be provided. If we are in fact creatures of a programme which we do not ourselves write, then any notion of the self as the unique entity which is ultimately efficacious with regard to action, seems to lose all credence. The clock, for instance doesn't contain some component over and above the material parts functioning according to natural laws of cause and effect and the prospect which (I, at any rate, think) would be presented by the truth of universal physical causation, is that persons are, in the significant sense, absolutely no different.

They may just be considerably more sophisticated in detail, that's all (and, as I will cite later, Dennett is one who mistakenly, I believe, holds the disparity in sophistication to be that which is ultimately important). Indeed by the end of the first paragraph of "One Determinism", Honderich has concluded that, if the determinism he is suggesting is, in fact true,

> It follows too that, we are not responsible for our actions, and, what is most fundamental, that we do not possess selves of a certain character.

The promptness alone of Honderich here, in floating this apparently calamitous conclusion, makes the apparent ignorance and dismissiveness of the likes of Austin and Strawson as regards determinism all the more unlikely. I hope that I have demonstrated here, that despite Austin and Strawson, the recounting of a few of what I think are now little more than banalities, reveals a real problem, and with all the monumental possible consequences it is felt, on the most primitive level, to have. The primary reason that I am impressed by the formalisation of the problem which I am about to consider, i.e. Carl Ginet's 'Might We Have No Choice?' is that it is a coherent and carefully thought articulation of those deeply and unmistakeably felt worries, which are instantly recognisable as such within the formulation. It is merely a postulated determinism i.e. Ginet is not especially arguing for its truth, nor is he even strongly suggesting that it is likely to be true. He is, however, arguing for its <u>logical</u> tenability, as well as claiming that we do not now know it to be false. And, as Ginet says, if the attempt is successful, that should be of sufficient interest in itself, and if it is not, then perhaps it will have been done in such a way as to make it possible to see why it has failed, and possibly even to see why any attempt at this task must fail. Austin and Strawson should perhaps note that it is not stated in one sentence; perhaps they made the error of expecting it to be, concluding that, since little of any interesting or useful content can be said about free will/determinism in one sentence, then there is no coherent thesis of determinism. Even Honderich, as I've said, takes a full

paragraph to state it, and also requires three separate premises, which he argues for in turn. This, as I've said, seems like most philosophical problems of any interest, and the likes of Austin and Strawson would do well to always bear in mind Charles Taylor's comments about philosophy in general, which I noted earlier.

Notes

1. In Essays on Freedom of Action, ed. Honderich.

2. In "Ifs and Cans", Proceedings of the British Academy, 1956.

3. In "Freedom And Resentment", Proceedings of the British Academy, 1962

4. Reprinted in Free Will, ed. Gary Watson.

5. In Brainstorms

6. In Freedom and Determinism, ed. Lehrer.

7. In "On Giving Libertarians What They Say They Want", Brainstorms.

8. In Freedom and Determinism, ed. Lehrer.

9. In Essays on Freedom of Action, ed. Honderich.

10. In "The Conceivability of Mechanism", Free Will, ed. Gary Watson.

2. A FORMALISATION

The initial stage of Ginet's enterprise involves formalising a certain relation between kinds of events (or states of affairs or circumstances), which he calls the relation of contingent necessitation. If 'A' and 'B' are descriptions specifying kinds of events (states of affairs, circumstances), then 'A' contingently necessitates 'B' if and only if

(1) 'A' does not entail 'B'

(2) every instance of 'A' is accompanied by an instance of 'B' (in a manner indicated in the descriptions 'A' and 'B'),

and

(3) no one ever has a choice as to whether or not an instance of 'A' shall be accompanied by an instance of 'B'.

Ginet cites cases where connections of contingent necessitation are already reasonably believed to hold e.g. if a piece of copper surrounded by ordinary atmospheric conditions becomes hotter, then it must expand or (say) if a small material body such as a bowling ball, is left near but not in contact with a very large material body, such as the earth, without anything between them but air, then the small body must move into contact with the large one. We can, notes Ginet, check these examples against the three defining conditions of contingent necessitation:

(1) We do in each case understand the antecedent and consequent in such a way that the former does not entail the latter, supposing that the terms "hotter" (in the first case), and the notion of gravitational force (in the second case) are both understood in their everyday, pre-theoretical sense.

(2) We do in each case believe, that every instance of the antecedent description (so understood) has been and will be accompanied by an instance of the consequent description.

(3) We do believe that no one ever has any choice as to whether or not an instance satisfying the antecedent description shall be accompanied by an instance satisfying the consequent description.

It is, I think, worth pausing and discussing some questions arising here. Note that, whilst Ginet's explication of 'contingent necessitation' is certainly rigorous, contingent necessitation seems, on examination, to add up to little more than our fairly basic, unreflective idea of straightforward causation. This is, I think, significant. And what is significant is that whilst it certainly could become important within another stage of the free will/determinism discussion, it is not, I think, at the moment necessary to attempt to settle time-honoured problems regarding the whole concept of causality itself. At the moment, we can, I think, remain entirely neutral as regards the questions about causality which the likes of Kant and Hume famously attempted to answer; we needn't worry, for the time being, whether causality is, for instance, an a priori condition of any experience of the world, or what sense we can make of the idea of something over and above the mere observation of constant conjunction. Whilst it will undergo important clarification throughout the course of this thesis, all we need be clear about for the moment is that <u>whatever ordinary causality is</u>, that is exactly what is being talked about here, in the concept of contingent necessitation. This feature is reinforced by the banality of Ginet's examples, i.e. the expansion of the heated copper, and the effect of gravity upon the bowling ball - whatever straightforward causation is exactly, that is what is going on here, and it is this, nothing more and nothing less which, largely, generates the free will problem. In the same way in which I can look out of the window to see if my father is arriving, and confidently settle this question without first having to settle the question of what persons are, what constitutes identity of persons over time, and whether realism or idealism about the physical world is true, we do not first have to settle all problems about causality to know that causation generates a problem about free will. Indeed,

in "One Determinism", Honderich makes this point:

> In what follows, incidentally, I shall speak in an ordinary way of causes. That is, a cause will be taken to be a chosen member of some set of conditions sufficient to produce an event or a state. To say of a state that it had a cause is to imply rather than assert, that the state was the effect of some sufficient set of conditions. It will make no substantial difference if one understands sufficiency in terms of constant conjunction or in terms of some stronger notion. (emphasis mine)

In fact, before returning to Ginet's specific offering, I feel it of some value to say some more on this and related issues. The primary reason I feel it important to do some clearing up here is that some have thought that until we have a complete understanding of the concept of causality and the closely related concept of a '"law of nature" (which we have already seen appear in Van Inwagen's formulation of incompatibilism), then we can barely make any sense of the thesis of determinism, let alone begin to enquire whether it is true. Others have thought that we know enough about those kind of concepts to know that their nature makes any attempt at forumlating the thesis of determinism collapse into incoherence (included in this class of concepts also, is the concept of "sufficient conditions"). Indeed, particular post-modern philosophical trends have tended to lend weight to scepticism about the import of concepts such as causation, sufficient conditions, and laws of nature. I feel it important to make some sort of response to these gestures, though (as will be borne out), I cannot begin to attempt an exhaustive analysis of the concepts in question. All I can do is finish what I have begun in the last few paragraphs, which is really to put the required concepts, and our understanding of them (or its lack) into some sort of perspective, vis à vis determinism. I said a moment ago that clarification would be required regarding the centrality of ordinary causality to the free will problem, and I will begin by saying that causation alone does not generate the free will problem. What generates a problem about free will is laws ("laws of nature", as we have seen Van Inwagen

call them), and despite Davidson's claim that causality is nomological in character, singular cases of causation are not qua cases of causation, an index to some or other law of nature; this view is supported by Searle in his book "Intentionality" and I will give much greater substance to this issue later, when I come to discuss the mind (compatibilist) argument. We will also see Ginet talk later of laws which would be required by his determinist hypothesis ("physiological-cum-environmental"). Indeed, Van Inwagen's distinction between the thesis of determinism and the Principle of Universal Causation will become very important later, not least when agent-causation gets onto the agenda. Determinism being defined by Van Inwagen as the thesis that there is at any instant exactly one physically possible future, he notes that in order to deduce determinism from the Principle of Universal Causation, we should ~~note~~ need at least three premises:

(1) if an event (or fact, change, state of affairs, or what have you) has a cause, then its cause is always itself an event (or what have you), and never a substance or continuant, such as a man (the truth of "agent-causation" would falsify this);

(2) if an event (or what have you) 'A' was the cause of an event 'B', then it follows, given that 'A' happened and given the laws of nature, that 'A' "causally necessitated" 'B', that 'B' could not have failed to happen.

(3) every chain of causes that has no earliest member is such that, for every time 't', some event in that chain happens earlier than 't'.

We can see from this immediately that as well as the concept of causation, in order to "get to" determinism, Van Inwagen has had to invoke the concept "law of nature". Without the latter concept, according to Van Inwagen, it may well be the case that while every event has a prior cause, the past nevertheless does not determine a unique future. The first thing I think it essential to stress is that we undoubtedly have the concepts of causation and law of

nature (and "sufficient condition"). I have just cited how little Honderich feels it necessary to say on these issues, and as far as causation is concerned, Ginet feels it unnecessary to say much more than to define "contingent necessitation" as he has done, and provide analogies, such as that of the expanding heated copper. When he gets to the flesh of his 'H' hypotheses, he feels that he can get away with talking straight off, of ("physiological-cum-environmental") laws, i.e. without really saying anything about the entire concept of a natural law. I've already indicated that I can really find little fault myself with these presuppositions of Ginet and Honderich. Going back in particular to Ginet's analogy, it seems to me that we confidently employ the required concepts to say that the expansion of the copper was caused by its being heated, and that this in turn can be explained by some or the other law ("of nature", if one wishes). It seems indisputable that we have no choice about what these laws are (it was not up to us whether the copper expanded on its being heated), and the worry which Ginet, Honderich or Van Inwagen seem to be floating for us, is that our own behaviour may in principle be explicable in exactly the same fashion, i.e. by being subsumed under laws which are not up to us. As I've indicated, it does not seem to me that any gaps in our precise understanding of causation and natural law (or any allegations about subject-input), undermine the coherence or reality of this worry. Van Inwagen, like Honderich, pretty much refuses to set foot in the minefield of causation but he does, I think, have some more than useful comments to make on the notion of a law of nature, comments which I think help defend the concept against its disparagers, and consequently, help defend the thesis of determinism against the charge of incoherence. He admits, right at the outset, that he does not know how to define "law of nature". However, he does suggest some constraints on an adequate definition of the concept and some necessary conditions for its application. For instance,

(i) the phrase 'is a law of nature' is a real predicate: it is typically and properly used in ascribing a certain property

to certain objects (unlike, say, 'exists', according to Kant, or 'is good', according to R.M. Hare),

(ii) the objects that have this property are sentences or propositions (non-linguistic entities expressed by sentences) or whatever it is that are bearers of truth-value: anything that is a law is also either true or false,

(iii) whether a proposition or sentence is a law is independent of what scientists or others happen to believe or happen to have discovered: a proposition, if it is a law, is unchangeably and objectively so, whatever mathematicians or others happen to believe or happen to have proved.

I think these remarks of Van Inwagen are of great importance, and I hope their importance will become clear as I proceed. The crucial point, which I feel cannot be over-stressed, is that, however much some may point to arbitrary or subjective elements in our formulation of natural laws, the fact remains that the alleged laws are true or false, only as a consequence of features independent of us and which we discover. Even if there are arbitrary or subject-dependent elements within the vocabulary which we use to approach an investigation, there are, within that vocabulary, objective facts. As Kenny says in "Freedom, Spontaneity and Indifference"[1] (when discussing Davidson), even if there is some sense in which laws are "linguistic", there is surely also an equally important sense in which they are natural, e.g. it is surely a natural impossibility that I cannot fly. Indeed, it seems to me that proper understanding of what I have just said leads one to see that there is something suspect about the whole linguistic/natural distinction. And it seems to be that when we entertain our half-formed, poorly conceptualised worry about determinism and free will, we are actually worrying that it may be a natural impossibility that I can act other than how I do act. As Van Inwagen says, 'law of nature' seems to be an intelligible concept and one we can't get along without if we wish to give a complete description of the world. Indeed, I think it worth quoting an

example Van Inwagen gives of this, and I think it is helpful that it is quite independent of any problems about determinism:

> I have recently read an article on the possibility of intersidereal travel in which the authors divide the unpleasant necessities of this sort into two categories: those imposed upon the travellers by the ignorance of the designer of their vehicle, and those imposed upon the travellers by the laws of nature....certain disadvantages of intersidereal travel are not going to be removed by technological advance as the corresponding disadvantages of inter-continental travel were removed. Inter-continental travel, now a matter of hours was once a matter of months or years. But intersidereal travel, if it should ever come to pass, will always be a matter of years or centuries. No technological advance could ever change this unfortunate fact for it is a consequence of the laws of nature.[2]

Whilst the concept of 'law of nature' may remain undefined, this kind of example surely demonstrates that it is an undefined concept we at least <u>have</u>. This, I hope, clears the ground, for much of what is to come, and I can now get back specifically to Ginet.

Ginet's next formal manoeuvre, which will be seen to have major ramifications as the discussion proceeds, is the notion of a <u>complete first-level description</u> of a person's behaviour during a certain period. This is a description that contains only the specification of all the places occupied by all the externally observable parts of a person's body throughout that period, relative to each other and to the adjacent environment, as well as the specification of all the sounds emitted by his body during that period. And we should really prick up our ears at this point, and take very careful note, that is, of what a complete first-level description does <u>not</u> include. The significant point regarding a complete first-level description is that it does not contain ascriptions of intentionality to persons, i.e. intentional verbs are strictly precluded. Within a complete first-level description, a person will not be described in the vocabulary of <u>action</u>; he will

not want, wish, believe, hope, desire, fear, intend, do, etc. By way of example, Ginet notes that "He repeatedly buzzed the other office" is not a first-level, but a higher-level description. Such a description cannot be a first-level one, since it contains an action-description; the person is described as doing something (i.e. buzzing the other office), which a first-level description cannot, by stipulation, accommodate. Ginet clarifies by going on to note that the statement "His arm and hand repeatedly moved so as to press his finger against the button on the side of the desk", does not ascribe any action to the person, and can therefore be a part of a first-level description. However, it still cannot be a complete one, since it makes no mention of what the rest of his body did at the same time. The significance of this notion, i.e. the complete first-level description, both what it includes and what it does not include, cannot be over-stressed, and should become clear as the discussion proceeds, since several attempts of some note at squaring free will with determinism have turned upon the (alleged) dualism of physical and intentional, the latter's non-reducibility to the former, and its explanatory indispensability. Indeed I will, in the course of the discussion, be forced to confront and examine various important claims in this regard, i.e. what the claim of the irreducibility of the intentional actually is, whether it is a justified claim, whether the intentional idiom is actually indispensable for a complete account of the world, and what import the answers to all of these questions have (if any) for human freedom. This project will, in fact, entail the examination of a specific attempt at demonstrating that there is no conclusive ground for believing that the intentional idiom is indispensable, this being the well-known Eliminative Materialism put forward by Paul Churchland in his essay "Eliminative Materialism and the Propositional Attitudes".[3] It is worthwhile, I think, to keep a note of these broad issues in mental brackets for the time being, until I actually rigorously examine them. Now I will press ahead with an elucidation of Ginet.

Having elucidated the notions of contingent necessiation and complete first-level description, Ginet is now in a position to

state his major hypothesis, which he calls 'H':

(H) Every temporal segment of every human being's behaviour has a complete first-level description 'B', and a series of antecedent sets of circumstances having the descriptions 'A_1', 'A_2',...'A_n' such that

(1) 'A_1' does not entail 'B',

(2) A_1 contingently necessitates A_2, A_2 contingently necessitates A_3,..., A_{n-1} contingently necessitates A_n, A_n contingently necessitates B,

and

(3) the human being in question clearly had no choice whether or not the antecedent instance of A_1 would occur.

This hypothesis entails that no human being ever has a choice as to whether or not his behaviour shall satisfy the first-level description it does. And it is impossible to over-stress the significance of Ginet's next claim, namely that this hypothesis entails that no one has any choice as to what descriptions of any kind his behaviour satisfies. Whilst this claim is far from gratuitous and has to be upheld in the face of some distinguished opposition, I feel it worthwhile, before I go on, to consider some opposition, to document another incompatibilist argument, which is somewhat germane, this being Van Inwagen's First Formal Argument. A look at this just now will, I think, help reinforce what could usefully be called the trans-vocabulary consequences of H (to which we have just seen Ginet refer). This will be seen, I think, to be especially useful, when I come soon to consider the 'language strata' theorists, who think that we can make sense of the idea of free will in the face of determinism through shifting vocabulary, i.e. from the physical (determinsim) to the intentional (free will). According to this, Ginet's claim that the truth of H would preclude any sort of choice as to what description our behaviour satisfies is false, since the truth of determinism need not be incompatible with our having choice about which action-description

our behaviour satisfies. As I say, I will soon consider (and attempt to refute) this position, but before doing so, let us look at Van Inwagen's First Formal (incompatibilist) Argument. As with Ginet's comments, Van Inwagen's argument here leaves it looking very difficult to make any sort of sense of the idea of free will in the face of exhaustive physical explanation, so before considering an effort at doing so in the shape of the 'language strata' approach, I will document it.

Some of the vocabulary of the argument will have to be simply assumed. Van Inwagen makes rather arduous attempts at cashing it all out, but I would prefer for the time being, at any rate, to simply assume that (as with concepts such as causation and 'law of nature') we are always dealing with concepts which we already have, and which admit of some sort of satisfactory account, even if we are not, in fact, presently in possession of such (though I will, in fact, be saying much more about causality, for one thing, later). 'To' denotes some arbitrary chosen instant of time earlier than J's birth, 'Po' denotes a proposition that expresses the state of the world at To, 'P' a proposition that expresses the state of the world at J, and 'L' the conjunction into a single proposition of all the laws of nature. The argument consists of seven propositions, the seventh of which follows from the first six:

(1) If determinism is true, then the conjunction of Po and L entails P.

(2) It is not possible that J have raised his hand at T and P be true.

(3) If (2) is true, then if J could have raised his hand at T, J could have rendered P false.

(4) If J could have rendered P false, and if the conjunction of Po and L entails P, then J could have rendered the conjunction of Po and L false.

(5) If J could have rendered the conjunction of Po and L false, then J could have rendered L false.

(6) J could not have rendered L false.

(7) If determinism is true, J could not have raised his hand at T.

That (7) follows from (1) - (6) is easy enough to establish, and I won't really concern myself with Van Inwagen's rather laboured and not obviously essential defence of each premise which follows what I have just presented as his argument. I will asume, for present purposes, that the argument is solid. It is, as I've said, more than useful to keep it established in our minds, as I move on to the 'language strata' theorists, such as Davidson. It's worth noting also that, unlike Ginet, who at least takes trouble to disclaim it, Van Inwagen doesn't even make any reference at this point to the 'language strata' defence of free will. He perhaps regards it unnecessary to even state the trans-vocabulary consequences of determinism. (Incidentally, Van Inwagen also, in an essay "The Incompatibility of Free Will and Determinism"[4], expresses admiration for Ginet's essay.) However, for the likes of Davidson, whilst there may, at the bedrock level of human behaviour, (e.g. neurophysiological and chemical) be laws of nature, expressible in a physical vocabulary, our freedom is nevertheless secured via the realm of the intentional, i.e. whilst a straightforward physical determinism with regard to human behaviour, such as Ginet's 'H' hypothesis, may well be true, human freedom is nevertheless unimpeded, since the mental enjoys an essential autonomy vis à vis the physical, mental concepts not being reducible to or translatable into physical ones and mental events being only contingently related to the physical events with which they are identical or correlated. I am myself ultimately in agreement with Ginet, that the truth of a hypothesis such as 'H' leaves us unable to sustain any attempt at locating human freedom within some other non-physical, allegedly autonomous vocabulary (and that Van Inwagen's First Formal Argument is germane to the point). However, those who have endeavoured to make this attempt at safeguarding freedom in the face of strict

physical determinism coherent and defensible, include Donald Davidson, Jerry Fodor, Jaegwon Kim and Daniel Dennett. The arguments have to be taken seriously, and before returning to Ginet, I will have to defend his claim regarding the consequences of the truth of 'H' against the opposition provided by the likes of Davidson, Fodor, Kim and Dennett. I now turn to this opposition, and I will call its propounders the 'language strata' theorists.

Notes

1. Reprinted in Honderich's Essays on Freedom of Action.

2. In Essay on Free Will, p7.

3 In Journal of Philosophy, 1981.

4. In Free Will, ed. Watson

3. THE 'LANGUAGE STRATA' THEORISTS

By far the most distinguished proponent of this position is Donald Davidson. Davidson wants to say that something significantly like Ginet's 'H' is true, but that recourse to a different (i.e. mental) vocabulary will reveal to us a refuge for our freedom. Contributors such as Jerry Fodor and Jaegwon Kim have attempted to make a stance such as Davidson's both clearer and more persuasive, and so Davidson's efforts, whilst I will find them ultimately defective, have to be taken seriously.

By way of preliminary, I think it interesting and useful (as often), to note the contextualising of Davidson on this point, which is provided by Van Inwagen. Van Inwagen is quite right, I think, to say, as he does at the beginning, that discussions of the question whether free will is compatible with determinism are usually not on a very high level. As he says, they are very often the work of compatibilists, and consist to a large degree in the ascription of some childish fallacy or other to incompatibilists (conflation of "descriptive" and "prescriptive" laws; failure to distinguish between causal necessity and compulsion; equation of freedom and mere randomness). As an incompatibilist, I have myself been frequently irritated by these asciptions, and Davidson places himself in this tradition when he writes:

> I shall not be directly concerned with [arguments for the incompatibility of freedom and causal determination] since I know of none that is more than superficially plausible. Hobbes, Locke, Hume, Moore, Schlick, Ayer, Stevenson, and a host of others have done what can be done, or ought ever to have been needed, to remove the confusions that can make determinism seem to oppose freedom.[1]

I can only hope that, as I go on, I will demonstrate that not only am I not suffering from any incompatibilist illusion that may exist, but that, despite the efforts of the writers Davidson lists, as well as Davidson's own dismissiveness, incompatibilism is indeed

correct. A large part of this demonstration will, of course, entail addressing myself to the compatibilist claims put forward by Davidson and the like. (Indeed, I hope that by the end of this work, it will appear reasonable to add the names of Lehrer, Dennett and Mackie to this list of somewhat self-assured compatibilists.)

It ought perhaps to be said, at the outset, that not everyone is likely to agree (immediately at any rate) that Davidson's position actually amounts to the purported truth of something significantly like Ginet's H, with which he must try to square freedom. However, I hope to make it clear why it seems to me that, at bottom, Davidson is a fairly ordinary physical determinist, and has a job on his hands making any sense of the notion of human freedom. Ginet, remember has said that the complete absence of choice regarding which complete first-level description is going to be true of us, undercuts any sensible idea of choice or freedom. This is exactly the claim which 'language strata' theorists dispute, and the key to understanding Davidson's position lies in unpacking this statement, which he makes in the aforementioned "Freedom to Act":

> ...His action, <u>in the sense in which action depends on intentionality</u>, occurs or not as he wills; what he does <u>in the broader sense</u>, may occur whether or not he wills it (emphasis mine).

In order to make some sense of this it is, I think, necessary to document more general features of Davidson's approach to mind and action. Davidson, as is well-known, is the foremost proponent of the theory Anomalous Monism. The theory is substance monist but property dualist, i.e. there is one type of stuff, namely physical stuff, but two types of properties (or property-<u>types</u>), physical and mental, neither being reducible to or translatable into the other. All events are physical, but some physical events are also mental events, and the properties which make these events mental are <u>irreducibly</u> mental in character, being conceptually tied to an exclusively mental idiom. The theory is <u>anomalous</u> in virtue of the

alleged Anomalism of the Mental, i.e. the assumption that there are no strict deterministic laws on the basis of which mental events can be predicted and explained. (In fact, whilst it isn't necessary to be particularly concerned with it at the moment, Kim argues that, despite Davidson's lack of rigour and clarity on the issue, the principle of the Anomalism of the Mental is actually the conjunction of two separate, more narrow premises, which Kim calls Psychophysical Anomalism and Psychological Anomalism. The first is that there are no laws connecting mental and physical phenomena, and the second is that there are no laws connecting psychological events with other psychological events, which can be used to predict and explain these events[2].) It is more than worth re-stating at this point that it is only our seemingly indelible conviction that our mental lives are both in some sense autonomous and ultimately efficacious with regard to our actions, which gives rise to the primitive, preverbal feeling of personal authorship (agency) and consequently the whole problem of free will. It is the prospect that this belief is entirely ungrounded that we are worried the sciences of man could eventually reveal to us. I re-state this, since Davidson is quite explicit in holding fast to these fundamental cornerstones of our belief in our own agency, i.e. the autonomy and the efficacy of the mental. And it should not, I think, be difficult to see that if, as I have suggested, Davidson's anomalous monism does bear significant similarities to Ginet's 'H' hypothesis, then Davidson is faced with difficult demands in making sense of psychological autonomy and efficacy. I have already touched on Davidson's thesis of mental autonomy: it is not obvious and perhaps not of special importance, how essential either or both of Psychophysical Anomalism or Psychological Anomalism are to this. What is of importance is the complete irreducibility of mental categories and their holistic character. Mental concepts cannot be reduced or translated into any other kinds, including the physical; the constitutive criteria of the mental, are wholly different to those of the physical, the former being constituted by its character as a system, i.e. in Davidson's words:

> Any effort at increasing the accuracy and power of a theory of

behaviour forces us to bring more and more of the whole system of the agent's beliefs and motives directly into account. (From "Psychology as Philosophy"[3]) (Kim calls this the Rationality Maximisation principle).

And as regards mental efficacy, Davidson wants to claim that rationalisations are causes of action. In his essay "Actions, Reasons and Causes", Davidson defends this claim against objections which it has had to face. I won't concern myself with this at the moment. What I will concern myself with is Davidson's attempt at coping with what appear, prima facie, to be four imcompatible strands within his position, i.e. rationalisations are causes of action, all events are physical, the Anomalism of the Mental, and the Nomological Character of Causality, this last one being that where there is causality there must be a law: events related as cause and effect fall under strict deterministic laws. (This last assumption is by no means embraced by everyone, as I've said.) I hope to show here that Davidson's attempt to retain psychological efficacy in the face of his anomalous monism, collapses in fact, into a determinism which seems to leave no more space for personal authorship than would Ginet's H hypothesis, were it true.

In tune generally, with his Psychophysical Anomalism and Psychological Anomalism, Davidson claims throughout that the mental is not the right kind of description to appear in deterministic laws, and this seems to leave him with an obvious problem: any singular statement of causality attributing the character of cause to a reason must be "covered" by some strict deterministic law, but if reasons are not the right kind of thing to get "into" deterministic laws, then how can singular statements of causality be so covered? Enter Davidson's assumption that all events are physical: reasons have some true description in a purely physical vocabulary, and it is as such that a singular statement of causality giving reason as cause is covered by a strict deterministic law. As Davidson puts it in "Actions, Reasons and Causes"[4],

The laws whose existence is required if reasons are causes of

actions do not, we may be sure, deal in the concepts in which rationalisations must deal...it does not follow that there is any law connecting events classified as reasons with events classified as actions - the classifications may even be neurological, chemical or physical.

As I've already said, both Jerry Fodor in "Special Sciences"[5] and Jaegwon Kim attempt to lay down a solid bedrock for a schemata such as that of Davidson. Fodor's paper is a highly general effort, concerning generally what he calls the "special sciences" (e.g. psychology, sociology) and in particular, how singular statements of causality within them can be true in the light of the apparent absence of strict deterministic laws within the special sciences, the apparent absence of type-identities between the categories of the special sciences and those of the physical sciences, and the apparently nomological character of causality. Fodor attempts to make projects such as Davidson's, i.e. of a "descent" from singular statements of causality within the special sciences to strict deterministic law within the physical sciences, seem both intelligible and likely. I will not concern myself with this particularly, since despite the admirably rigorous approach of Fodor, I don't think that anything said by him at all alleviates the worries I will cite in connection with Davidson's approach. Kim's essay, on the other hand, is specifically designed to make Davidson's particular position more plausible and indeed, as I hope to make clear, the factors cited by both Kim and Fodor by way of making Anomalous Monism more persuasive are precisely the kind of factors which I think must ultimately prove its undoing. Indeed, as I hope to show, it is precisely the feature which particularly concerns us here, vis a vis free will, i.e. the efficacy of the mental, which Anomalous Monism seems to leave looking most vulnerable.

Kim is wanting to distinguish between the kinds of relationship between the mental and the physical which Davidson's anomalous monism can allow, and those which it can't. The point of departure for Kim is the entire concept of a law, and the crucial role that

considerations of lawlikeness play in Davidson's central argument. Kim asks us to consider a domain U of objects, and two sets, F and G of properties, where U is a set of medium-sized material bodies, F a set of colours and G a set of shapes. We would not expect to find regular correlations between colours and shapes; we would not expect true generalisations of the form (using Kim's notation):

(A) Every object in U with colour C has shape S,

or of the form

(B) Every object in U with shape S has colour C.

However, we may in fact find, say,

(C) Every red object in U is round.

The significant point, though, is that given what we know about colours and shapes, we would not take the truth of (C) as indicating a lawlike connection between being red and being round; the truth of (C) is a coincidence, not a matter of law. Kim suggests, similarly, that we let the domain be the set of persons, and F and G respectively the set of psychological properties and the set of physical properties. Davidson's point is that even if we should find a true generalisation of the form

(D) All persons with mental property M have physical property P,

we will not, and should not, consider this a law.

For Davidson, a law is distinguished from a mere generalisation by two marks:

(1) it can support counterfactuals and subjunctives, and

(2) it is confirmable by observation of its instances.

(C), for instance, meets neither of these criteria; it fails to back a counterfactual such as (Kim's example) 'If bananas were red, they would be round', and the only way it could be confirmed is by an exhaustive examination of all the objects in the domain. Davidson's arugment is designed to show, not that there can be no psychophysical generalities of the form (D), but that there can be no psychophysical laws. Davidson thinks we can show from the very idea of what it is to be psychological, that no generalities of the form (D), whether true or false, can be lawlike. In order for Davidson's arguement to seem at all cogent, we must identify the features of the mental and those of the phsyical which, whilst apparently tolerating true psychophysical generalisations, are inimical to these generalisations being lawlike. I have already touched on this, but I think that it is more than worth some re-statement and expansion. Davidson claims that both the mental and the phsyical are characterised by certain 'synthetic a priori' laws, which are constitutive of our conception of them. In the case of the physical, these constitutive criteria make possible the formulation of precise laws. Among these criteria, for instance, are principles that make physical measurement possible, such as, the transitivity of 'longer than' or 'earlier than'. According to Davidson however, the constitutive criteria of the mental are quite different, including the likes of coherence, rationality, and consistency (Rationality Maximisation) and indeed, we can extract from Davidson's seminal and most famous essay "Mental Events"[6], that

> there are no strict psychophysical laws because of the disparate commitments of the mental and physical schemes,

and that

> there cannot be tight connections between the realms if each is to retain allegiance to its proper source of evidence.

Several however, including (as will be echoed a bit later) Honderich, have wondered about this, i.e. <u>how</u> the presence of nomological links is inconsistent with each system's retaining its

allegiance to its constitutive principles. And the crucial move on Davidson's part which we must make sense of is the claim that the essential characteristic of the mental, namely rationality, must clash with an essential feature (or at least, a constitutive element) of the physical, namely the absence of rationality. It is this apparent clash which Davidson seems to see as precluding psychophysical laws.

This seems, at the very least, however, a curious claim on the part of Davidson. Aren't our physical theories also coherent, rational and consistent? If a theory in the physical sciences appears incoherent, irrational or inconsistent, don't we therefore abandon it with as little hesitation, as in the case of a purported explanation of someone's behaviour?

Kim seems to be well aware of the gaps needing filling here, and I will now recount and consider his attempt at this. Kim is attempting to show why the claims Davidson wants to retain on behalf of the mental, i.e. (primarily) allegiance to its "proper source of evidence" and efficacy with regard to action, could sustain contingent psychophysical generalisations, but not psychophysical laws. In the face of physical determinism (which Davidson is, at bottom, committed to), I will be arguing ultimately that both contingent psychophysical relations such as Davidson, Kim and Fodor think there are, and nomological psychophysical links (such as Honderich thinks there has to be), present very serious obstacles for any coherent notion of freedom. I will be claiming that both notions put an intolerable strain on the notion of psychological efficacy and so, if deterministic law can indeed by discovered at the neural or chemical level (as Davidson, Kim and Fodor think likely, and as Ginet's 'H' hypothesis postulates), then a refuge for freedom cannot be found through recourse to the safe and soothing irreducible vocabulary of the mental, i.e. I will be arguing that Ginet is correct in his claim that the absence of choice with regard to what complete first-level description I satisfy entails what I have already called a "trans-vocabulary" absence of choice. However, this should become clearer after examining Kim's attempted

patching-up of Davidson.

We are to allow that there are neural states N_1 and N_2, which are nomologically coextensive with M_1 and M_2 respectively, where M_1 and M_2 are mental states; that is, we have laws affirming that, as a matter of law, N_1 occurs to an organism at a time just in case M_1 occurs to it at that time; similarly for N_2 and M_2. The neural states, N_1 and N_2, being theoretical states of physical theory, have conditions of attribution, that is, conditions under which their attribution to an organism is warranted. Kim stresses that what matters is only that the ascertaining of whether they hold in a given situation is regulated by the constitutive rules and principles of physical theory, not by those of the mental. If we let C_1 be an attribution condition for N_1 and C_2 an attribution condition for N_2, then we have:

(1) Necessarily, if C_1 obtains, N_1 occurs,

the psychophysical law

(2) Necessarily, ~~if~~ M_1 occurs if and only if N_1 occurs,

whence

(3) Necessarily, if C_1 obtains, M_1 occurs,

and in the same way

(4) Necessarily, if C_2 obains, M_2 occurs.

It is worth stressing that (3) and (4) affirm that, when a certain set of physical conditions hold, a specific mental state necessarily occurs, that we must attribute to an organism this mental state if those conditions are observed to obtain for it. But, as Kim remarks in "Psychophysical Laws":

...this means that the rationality maximisation principle as an

essential constraint on the attribution of mental states is in danger of being pre-empted, for the determination of whether these physical attribution conditions obtain is not subject to the constraint of this principle. (3) and (4)...would force this (rationality maximisation) rule to share its jurisdiction over mental attributions ...by becoming so intimately associated with C_1 and C_2 which are under the jurisdiction of physical theory and its constitutive principles, they have in effect ceased to be mental states (emphasis mine).

This, it should be clear, is a crucial notion. I agree with Kim here that the truth of something like (3) and (4) would have calamitous consequences for the supposed autonomy and irreducibility of the mental; the idea of the efficacy of the psychological qua psychological, of the purposive, would seem to have an unmanageable strain put upon it. As Malcolm comments in "The Conceivability of Mechanism", the neurophysiological cannot be more basic than the purposive, the latter cannot, qua purposive, turn out to be dependent on the former; there is an a priori relationship within the purposive, it is a closed system of conceptual truths, logically insupportable by experience (e.g. my desiring a beer is logically linked to my getting up and opening one), whilst links between neurophysiological states are contingent. It is perhaps in order to add some clarification of this point here, since it will be seen, I think, to be of greater importance as I proceed, particularly with regard to what kind of concepts can occupy the 'A' slots within Ginet's 'H' hypothesis.

It has, for instance, seemed to some that the links between (say) desire and action are causal and must therefore be contingent, this seeming to preclude my suggestion that Malcolm is right when he says that there is an a priori relationship within the purposive, it being a closed system of conceptual truths. However, I think that the belief that there is a conflict between those two claims rests on a confusion. I think it is possible to hold (as I would want to) that there is an a priori relationship within the purposive, that some intentional states are a cause of action and that there is

something contingent about the relationship between (say) desire and action. Let me explain. First, what is it precisely, that is a priori, conceptual, about the relationship within the purposive? The answer to this, I think, is discoverable, if we ask what the essential nature of a purposive state is. And in fact, I think that the answer to this points towards the solution to the whole problem. The essential nature of a purposive state links it to action, i.e. where there are no 'countervailing factors' (to use Malcolm's expression). For instance, from knowing that John desires a beer, we can reasonably conclude, straight off, a priori, that if there are no countervailing factors, then John will take a beer. Conversely, if someone told me that John desired a beer, and on my offering him one, he did not take it in the absence completely of any countervailing factors such as health considerations, later plans or whatever, then I would be justified in being highly sceptical of the report I received about John's desire. This is because it is part of the essential nature of a desire to move one to action in the absence of countervailing factors - there is a clear sense in which the link between purposive states and action is not accidental. My hoping that it doesn't rain tomorrow, is insufficient grounds for expecting me, straight off, from just knowing this fact alone, to do anything whatever. However, from knowing that I desire to go to the cinema tonight, it is entirely reasonable to expect me straight off, to do something. If I don't do anything at all, then it is reasonable for someone to look around for reasons why, for countervailing factors. And I don't think that it should now be too difficult to see how this needn't conflict with either of the suggestions, that purposive states cause actions and that there is something contingent about the relationship between the two. There is, as I have said, an essential link between purposive states and action: it is an essential part of the nature of a desire for the subject to move to bring about its satisfaction in the absence of countervailing factors. And there seems to me to be no reason why this essential relation between purposive states and actions cannot be causal. It seems to me entirely reasonable to suppose that it is essential to the nature of a desire to cause action in the absence of countervailing factors: what is wrong with

the suggestion, for instance, that it is essential to my desire to go to the cinema tonight, that in the absence of countervailing factors, it causes me to go to the cinema tonight? As I've already suggested, it seems to me to be difficult to make sense of it qua desire, otherwise. And it is (not surprisingly by now, I hope), in the notion of 'countervailing factors', that I think the contingency which seems to be a fact of the relationship between purposive states and action can be found. That is, a desire will only cause the action with which it is essentially linked if there are no countervailing factors e.g. John's desire for a beer will not cause him to take one if (say) he wants to study in an hour. (It is worth saying just now that countervailing factors are essentially factors which are causally relevant, and this notion will be explained shortly in connection with Honderich's comments on anomalous monism, and will be seen to be of some importance, as I go on.) The significance of countervailing factors drives a wedge of contingency between desire and action, but I hope to have demonstrated that this does not entail denial of an a priori relationship between them. The task, I think, is simply to identify where the a prioricity lies and where the contingency lies. Indeed, I hope that the general suggestion I am putting forward here will appear both clearer and more obviously defensible, when I come, much later on to document the contribution on these issues, made by John Searle, in Intentionality (specifically, the chapter entitled "Intentional Causation"). Searle, I think, makes a penetrating contribution, to clarifying precisely where the a priori or 'logical' element of the intention/action relation resides, and the general importance of this within the theory of action, and of course within problems surrounding free will, should be seen to be considerable. For the time being, at any rate, I will rest content with saying merely that I think the three intuitions I have mentioned, namely that there is an a priori relationship within the purposive, that some intentional states are a cause of action, and that there is something contingent about the relationship between purposive states and action, can be reconciled by saying that it is an essential part of the nature of purposive states to cause action in the absence of countervailing factors. This will, I hope, be seen to have some importance when I

return specifically to Ginet in the next chapter, particularly with regard to the question of what can occupy the 'A' slots, and (as I've just said) when I eventually come to consider Searle's contribution to the issues and their significance for questions within free will/determinism.

Getting back to Davidson and Kim, the truth of (3) and (4) would seem to have this consequence of compromise for the psychological, of making the neurophysiological more basic than the purposive, of making the a priori connections within the purposive contingent in fact upon the neurophysiological. The most basic question within free will/determinism (though by no means the only interesting question, as we shall see) is whether purposive principles have any application to the world - if they don't, then we can wrap the whole thing up now - but the truth of (3) and (4) would leave it very difficult to see how purposive principles have any genuine application to the world at all. Davidson is, of course, as I've already said, wanting to avoid this consequence; despite his belief that physical determinism is true at the bedrock level (à la Ginet's 'H' hypothesis) he still believes that he is able to hold on to full-blown mental efficacy through recourse to the distinct and irreducible categories of the mental. As I've already said, I don't think that Davidson can, in fact, sustain his claim of mental efficacy in the face of his ground-floor physical determinism (I will attempt to provide more reasons for this in a moment), but what is undoubtedly certain at least is that it could not be sustainted if (3) and (4) were true. This is because, if (3) and (4) were true, it would make it very difficult to see what the mental descriptions (M_1 and M_2) would actually add by way of explanation of anything. And this factor would be merely reinforced by the fact that according to Davidson, each mental event is identical with some physical event i.e. there would seem to be absoultely nothing lost by simply explaining actions by means of physical/neural causation - as Kim has said, the specifically mental, with all its irreducibility, efficacy and allegiance to appropriate source of evidence, would seem to just perish away, with desperate consequences for psychological efficacy and therefore, of course, human freedom. The mental would seem here at best, to be

relegated to the status of superfluous accompaniment to what goes on inexorably by means of physical causation, and so much the worse for our illusions of agency.

However, what of the Davidson alternative? Can psychological efficacy be sustained either in the face of merely contingent psychophysical identity? I think not. In an excellent article, "The Argument for Anomalous Monism"[7], Ted Honderich considers the consequences of the absence of nomological links between the mental and physical for the autonomy and efficacy of the mental. Honderich concerns himself, in particular, with the notion of a causally relevant property. By way of clarifying this notion, Honderich considers an example: some French green pears are placed on a scale, causing the pointer to move to the two-pound mark. Given that the pears being both green and French are essential properties of the event, are the pears being green and French causally relevant to the given effect? The answer, of course, is "no"; the pears could have been yellow and Italian, for instance, and the pointer would still have gone to the two-pound mark had the pears still weighed two pounds. And there will be a lawlike connection, it seems, between the pears' weighing two pounds and the pointer's going to the two-pound mark. In general (as Honderich points out), it follows from the fact that event E_1 caused event E_2 in virtue of a property f of E_1 and a property g of E_2 that E_1 and E_2 are in lawlike connection wholly in virtue of properties f and g. Honderich goes on to pose the crucial question: if a mental event causes a physical event (as Davidson would want to claim), what is the causally relevant property of the mental event? In order to preserve the efficacy of the mental, it seems (surely) that mental events must be causal as mental, i.e. the causally relevant properties must be mental. This would surely force us to strengthen Davidson's premise of the Nomological Character of Causality into one of the Nomological Character of Causally Relevant Properties and, in the face of mental-physical causation, we are, it seems, left with a denial (despite Kim) of Davidson's thesis that there are no psychophysical lawlike connections, i.e. a denial of Psychophysical Anomalism. To retain Psychophysical Anomalism would

leave the connection between the mental and physical accidental i.e. there would be no nomic necessity about the event as physical being the mental event it is. Therefore we have lost the efficacy of the mental. According to Davidson's Psychophysical Anomalism , it is nomologically inessential to the event's being the physical event it was that it was the mental event that it was. Hence, it is inessential to the relevant effect, the action, that the event was the mental event it was. I think that these criticisms made by Honderich are devastating, every bit as damaging as the consequences brought out (it seems) by Kim of the prospect of presence of nomological links between the mental and physical, i.e. when either is conjoined with physical determinism. It seems to me that, either way, the mental cannot win: in both the presence and the absence of nomological mental-physical links, physical determinism (despite the desperate attempts of Davidson, Fodor, Kim and others to locate it at some apparently inconsequential level) deprives the psychological of its autonomy and efficacy and makes it intractably difficult to sustain any notion of free will. Indeed, the stances adopted by Honderich in "The Argument for Anomalous Monism" and his essay "One Determinism" are both revealing in this respect and, I think, slightly curious. Let us dwell on Honderich for a moment. I have already documented his criticisms, vis à vis the efficacy of the mental, of the notion, prevalent in Davidson's Anomalous Monism, that there are contingent links between mental events and physical events with which they are (allegedly) identical. However, one should beware of leaping therefore, to the conclusion that Honderich is himself of the belief that the presence of nomological mental-physical links would make it easier to sustain a coherent notion of free will, if physical determinism were true. On the contrary, he seems to be in complete agreement with what I have just said, namely that, in the face of physical determinism, the presence of nomological psychophysical links would not help us one iota towards regarding ourselves as free agents. This seems clear from "One Determinism", where the determinism he actually argues is true, features nomological psychophysical links, and which, as he states explicitly, must leave us lacking free will. It is, I think, worth spelling this out in some detail.

Honderich's ("One") determinism is a conjunction of three fairly straightforward premises:

(i) States of the brain are effects, the effects of other physical states.

(ii) Many states of the brain are correlates of conscious states (The Correlation thesis).

(iii) States of the brain are causes, both of other states of the brain and also of certain movements of one's body.

The latter are actions.

The crucial premise, for my present purposes, is the second one, the Correlation thesis. Before being any more specific about it however, it is very useful to merely bear in mind that, from these three straightforward premises, Honderich concludes that, on every occasion when we act, we can only act as in fact we do, that we are not responsible for our actions and that we do not possess selves of a certain character. Both the first and third premises would get the assent of the likes of Davidson, Fodor and Kim (~~and of course, Ginet~~). They are, at ground-floor level, as I've already said, physical determinists, just like Honderich, and in fact, should anyone still think that Davidson and the like are being badly represented by being thought of as physical determinists (or indeed, determinists of any description), it is, I think, salutary to note that, in his essay "Rational Explanation of Actions and Psychological Determinism"[8] , David Pears takes up a position significantly similar to that of Davidson, and goes on to state that it would be fair to regard it as either a psychological or physical determinism. (Indeed, I will go on later to make the important claim that the only "determinism" which it really makes sense to speak of is physical determinism, a fact not only implicitly acknowledged by Davidson, but also haveing significant consequences throughout the family of debates on free will/determinism.) However, Honderich's second premise would clearly not get the

agreement of the anomalous monists. This is because the correlation thesis claims that there are much tighter (i.e. nomological) psychophysical links than the anomalous monists believe (as we have just seen Kim try to spell out) the mental can bear. The Correlation thesis is that

> Any particular description of consciousness D is true of an individual if and only if his brain or part of it is in one particular state or sequence of states S.

Before saying any more, it is therefore perhaps worth noting that Honderich realises that, rather than talk of states of a part or whole of the brain, we might be better, given certain propositions in contemporary neurophysiology, to talk of states of the central nervous system or a part of it. At any rate, Honderich intends the term part, to be no more than a stand-in for other more adequate descriptions. A word is also in order about what is meant, and what is not entailed, in the notion correlate: there is no implication here regarding the non-identity, or identity, of brain states and conscious states. All that is to be understood concerns descriptions of consciousness and brain states, and I agree with Honderich that the numerous problems concerning the individuation and truth conditions of descriptions of consciousness are not impediments to the thesis, and needn't be settled. Also the term brain state may be so construed as to take as its referent either one state or one disjunctive set of states.

The crucial point for Davidson and Kim, the factor which must prevent them embracing this Honderich premise is the iff; from seeing that someone's brain is in one of a particular disjunctive set of states, we can (in principle, at least) conclude that he is in a particular mental state. We have already seen Kim elaborate upon how such a prospect would undermine the autonomy and efficacy of the mental by making the neurophysiological more basic. I have already expressed agreement with this, and there can be absolutely no doubt from what Honderich has to say, that he is also in agreement that the presence of psychophysical nomological links

(such as he documents here) is as damning a prospect for mental efficacy (and therefore free will), as he has already claimed (in "The Argument For Anomalous Monism") the absence of nomological links is, in the face of physical determinism and mental-physical event identity. I already noted that he concludes that his three premises entail that on every occasion we act we can only act as in fact we do, that we are not responsible for our actions, and that we do not possess selves of a certain character. That is, despite his claiming, in "The Argument For Anomalous Monism", that the presence of merely contingent psychophysical links has intolerable consequences for the integrity of the mental (which Kim seems not to appreciate), he seems to be aware, (in "One Determinism", at any rate) also that the mental can do little better in the presence of nomological psychophysical links. I do myself believe (as Honderich seems to) that contingent or nomological, it's really all the same at the end of the day, i.e. the game is up, the mental simply cannot win, once physical causation is allowed to be ultimately so comprehensive as either Davidson, Kim, Fodor, Pears or Honderich (or Ginet of course) suggest it is. Once Honderich's first and third premises are admitted, it seems to me not to matter, from the point of view of mental efficacy and free will, whether what goes in to the slot waiting for the second premise, states contingent psychophysical relations, or supposedly "stronger" nomological ties. Kim is right to say that the mental cannot bear the weight of nomological mental-physical ties and Honderich is right if he is saying (as I think he is) that, in the presence of physical determinism, a la his first and third premises, the mental can bear neither the weight of nomological ties with the physical, nor that of merely contingent psychophysical links. Either way, nomological or contingent, it is surely the same, i.e. how do I get into things, how is my mental life qua mental life, choices qua choices, causally relevant? All of my actions would be the outcome of a set of sufficient (physical) conditions, so where do I, with my (seemingly) complex intentional network, comprising agonising choices, decisions, feelings, etc. get in there? There may be contingent psychophysical links, there may in fact be nomological links, but then so what? The significant point is surely that, no matter, my

actions are wholly explicable in terms of sets of sufficient (physical) conditions and indeed (in the presence of whatever links) why should we even be especially interested in what would appear to be the superfluous and unnecessarily complicating factor of mental facts? If all of my actions are wholly explicable in terms of the vocabulary of physical causal law, then what can possibly be gained by introducing the intentional into explanation of my actions? Physical causation will trundle on, no matter. However much sweat is broken at this point, there seems to be no clear way in which Ginet's claim (with which I started this section) looks over-hasty, i.e. if no human being ever has a choice as to whether or not his behaviour shall satisfy the first-level description it does, then no one has any choice as to what descriptions of any kind his behaviour satisfies. I hope that I have shown so far however that whilst it may be correct, it is certainly not gratuitous.

I have already mentioned Daniel Dennett (a functionalist) as among those who have been unconvinced that comprehensive physical causation of our actions deprives the intentional of its full-blown efficacy. And indeed, although I will, again, have to disagree with him, he cannot be ignored, not least of all because some of his attempts at appeasing those with free will predilections could be regarded as either novel or desperate or both. The Dennett enterprise is constituted in two essays, the well-known "Mechanism and Responsibility"[9] and the aforementioned "On Giving Libertarians What They Say They Want". Having looked at the likes of Davidson and Kim, the former essay appears as fairly mainstream compatibilism, whilst it is the latter where I think the trickery and hint of desperation makes itself apparent. Both essays attempt, of course, to downgrade the supposed impact that the prospect of comprehensive physical (causal) explanation of our actions must have on our conception of ourselves as free agents in the world. I will eventually look at the latter work, but for the time being, let us have a look at "Mechanism and Responsibility". This essay leans heavily on a purported analogy between human beings and the finite mechanical systems which comprise computers. Dennett is preoccupied here with the question of whether a mechanistic explanation of a set

of events, i.e. explanation in terms of straightforward physical causation, is incompatible with an intentional explanation of the same set of events. However, I hope that I have made it clear by now that the mere prospect of compatibility of physical and intentional explanation has little import for the issue with which I am concerned, i.e. free will. This point should be clear from the treatments alone which I gave Davidson's Anomalous Monism (or perhaps, it would be more accurate to say, Kim's re-statement of it), and Honderich's ("One") determinism. Indeed, it is worth saying yet again at this point, and indeed I feel it cannot be over-stressed, that Honderich's determinism clearly allows intentional explanation of actions, yet its unavoidable upshot is, as he says, that on every occasion when we act, we can only act as in fact we do, we are not responsible for our actions and that we do not possess selves of a certain character. His Correlation thesis quite obviously introduces the intentional lives of humans into the picture, and allows us to explain actions by invoking it in the fairly normal way. However, again I feel that I cannot over-stress that this prospect would hardly be enough to give the concept of freedom any genuine content. Indeed, a sceptic about free will may quite easily concede the compatibility of physical and intentional explanation, without being convinced one iota that there is such a thing as free will. What is needed surely, for free will is a specifically psychological efficacy, a psychological efficacy which can only accrue if the intentional is not only irreducible, but also entirely indispensable as a means of explanation. And that surely means that physical determinism cannot be true if free will claims are to be sustained. This is because, if physical determinism is true, then however much we may invoke the intentional ordinarily to explain things, we need not invoke the intentional ever, i.e. we could (in principle, at least) explain every action ever performed, every thought ever had, every statement ever made, wholly in the vocabulary of physical causes and physical effects. A look again at the three premises of Honderich's determinism should confirm this. Let me re-state the first and third:

(i) states of the brain are effects, the effects of other physical states,

(iii) states of the brain are causes, both of other states of the brain and also of certain movements of one's body.

The latter are actions.

It should not, I think, be difficult to appreciate that, if (i) and (iii) are true, then every action ever performed, every thought ever had and every statement ever made can be entirely accounted for without recourse to intentional vocabulary, without mentioning anyone's desires, beliefs, hopes, fears, intentions, purposes, projects, etc. We could (in principle, at least) explain all of these phenomena simply by talking of neurons, nerve cells, states of the central nervous system, etc. and whatever neurophysiological tools happen to be in vogue at the time. Of course, the Correlation thesis (the second premise) would allow us to invoke intentional concepts in explanation, would facilitate this but the crucial point would be that whilst we may, if only for sake of simplicity, use the intentional idiom in explanation, we need not. That is, the intentional would, in fact add absolutely nothing, it could be done away with entirely, and nothing would be any different. It would not be indispensable, it would not have an essential explanatory function, and this is surely what free will cannot sustain. For mental efficacy to genuinely explain, it must surely be the case that the actions of ours which we regard as free, those actions which are genuinely and full-bloodedly ours, can only be given a complete explanation through recourse, at some point at least, to the intentional idiom, to a vocabulary of purposes, intentions, hopes, fears, etc. For instance, my opening the fridge for a can of beer or my going to London to see an exhibition, can surely only be explained, if they are to be regarded as free actions of mine, by talking of my desire for a can of beer, and my belief that there was one in the fridge or, in the latter case, my belief that a certain exhibition was taking place in London and my belief (say) that I would enjoy it (of course, as I will discuss more specifically later, I am not claiming that the indispensability of intentional

explanation is sufficient for free action, but only necessary). The truth of a determinism such as that suggested by Honderich or Ginet would deprive the intentional of this indispensable position, kill the efficacy of the mental, and leave free will without a foothold. That is, despite the possibility of intentional explanation of my opening the fridge or going to London, the prospect of its being fully accounted for in the vocabulary of physical/neural causation would surely leave it impossible to see where I get onto the stage. That is why I find Ginet's clarity and explicitness on this point particularly helpful, and this should become clearer when I eventually return to him. The 'H' hypothesis is a straightforward physical determinism, little different to that of Honderich's ("One") determinism, but Ginet does not muddy the waters in any way by even considering what kind of relationship (if any) the intentional may have with the physical, and whether the intentional could even get a foothold in explanation. The complete first-level description contains purely physical concepts and the 'A' slots can only be occupied by physical concepts also. That is, Ginet's 'H' hypothesis is, I think, like Honderich's determinism without the Correlation thesis, i.e. it is really, I think, like a statement of Honderich's first and third premises, and any free will defenders cannot be temporarily deceived by the presence of something unnecessarily complicating such as the Correlation thesis. Perhaps it should be pointed out that Ginet is not particularly at pains to deny that something like the Correlation thesis may be true; for all he seems to be concerned, it may be true, but this surely points to what is really significant, i.e. from the point of view of free will, if H is true (or for that matter Honderich's determinism), it does not matter in the slightest whether the Correlation thesis is true. The truth of H, or anything significantly like it, surely damns free will, and as I've already suggested, this should become reinforced even further when I get back to Ginet.

As I've already said, equivalent worries for freedom must accrue even in the presence of a physical determinism so diluted and undercover as that of Davidson's, and of course, that of Dennett also. As I've already touched on, in "Mechanism and Responsibility", Dennett goes to some lengths and constructs analogies with both

computers and less sophisticated living creatures in defence of his claim that humans are finite mechanical objects. The point of both analogies is really little different: the bottom line is (as I've said) that the historical horror that the prospect of comprehensive physical explanation of human endeavour would rule out intentional explanation, is ill-founded. It should not be difficult to imagine the supposed import of the computers case: it is, says Dennett entirely in order to adopt both the mechanistic stance and the intentional stance towards computers. Everything they do can (with the required knowledge) be accounted for in terms of physical causation; however it is also, suggests Dennett, fair to talk of the activities of the computer in straightforwardly intentional terms, i.e. we can reasonably say that the computer is thinking, calculating, pondering, believing, etc. This, Dennett thinks, can be extended to "machines" even so complex and sophisticated as humans: like the computer, mechanism can wholly embrace us. Simultaneously however, we are (like the computer, it is alleged) also finite and rational - we can be characterised in intentional terms. Dennett does make the concession to traditional incompatibilism that, whilst it would in principle be possible, humans are such complex mechanical systems that it would be wildly optimistic to ever hope that we could achieve comprehensive mechanical explanation. And, with specific regard to the issue of free will and moral responsbility, Dennett concludes, in Section VII, that

> The Intentional stance towards human beings which is a precondition of any ascriptions of responsbility, may coexist with mechanistic explanations of their motions.

As the use of Dennetts alleged "computer" analogy wasn't difficult to imagine, I hope that it is not now difficult to imagine my response to it, vis à vis free will. Whilst it may not of itself be of ultimately crushing importance, I don't actually think myself that it is fair to say that computers, however sophisticated, have genuine intentional faculties - they don't actually think or believe anything, but at best, only undergo analogies or representations of

these procedures (it is unnecessary to pursue this specific point here, but precise comments on the matter, I think, can be found in John Searle's essay, "Analytic Philosophy and Mental Phenomena"[10]). However, the important point is that, even if they do, it would not, I think, be reasonable to hold them responsible for anything they think or believe, or calculate or whatever (in Strawson's language, to adopt the "personal reactive attitude" towards them[11]) and so, if humans are more complex finite mechanisms, and nothing besides, then there seems no more reason to adopt reactive attitudes towards humans than towards computers. Ironically, the case of the computer does turn out, I think, to be uncannily instructive. Even if it is fair to regard the computer as having a full-blooded intentional existence, there can be no doubt that there is a very clear, identifiable sense in which the physical is more basic than this; whatever the computer thinks, believes, calculates or whatever (whether one regards this as genuine or metaphorical intentionality) is entirely dependent on physical causes and effects, which is in turn dependent on how someone has constructed it. Intentional concepts, however usefully they may be utilised in explanation of the computer's movements, could be entirely dispensed with, leaving us with a complete explanation of everything the computer "says" in terms of physical causation. It is precisely because of this, I suggest, that the 'participant' attitude towards the computer is not a possibility. Similarly, regardless of the possibility or usefulness of intentional explanation, if humans are similar in this significant sense to computers, there is then a serious problem regarding rational justification of the participant attitude. If the neurophysiological is more basic than the intentional (and the equivalent is clearly so with computers), then it is not at all easy to see how the participant attitude can be justified. That is why it is extremely short-sighted of Dennett to say (as I've quoted earlier) that it is the mere possibility of the Intentional stance which is the precondition of any ascriptions of responsbility. As I hope to have made clear by now, the mere intentional stance is not itself enough to ground the participant attitude: what is also a condition is a belief about the mental life of the person in question, i.e. that it is, in some crucial sense, irreducible,

autonomous, indispensable to an explanation of their actions. It cannot refer us back to something more basic (e.g. the neurophysiological) in which it is grounded, and which could explain exhaustively anyway (in principle); as soon as something like Ginet's 'H' hypothesis, Honderich's first and third premises, or Dennett's "finite mechanisms" is admitted, then the participant attitude is in real trouble. This perhaps make it unsurprising that, in order to ground participant attitudes, Dennett is forced to fall back rather tamely on Strawsonian "irremoveable component of our humanity" theses (about which I will say more nearer the end of this thesis). Dennett's analogy with other creatures of the animal kingdom really does little better, I think. He places the (by all accounts, unintelligent) Sphex wasp at one end of a spectrum and humans at the other, this being the line of mechanical sophistication. Not surprisingly, the point for Dennett is that the wasp is so lacking in intelligence because of its ultra-primitive mechanical organisation, and humans have so much more intelligence because of their much more sophisticated mechanical organisations. At some apparently arbitrary point on this line of mechanical sophistication, we being to invoke a completely different mode of explanation, i.e. the intentional mode, which is (supposedly) running in tandem the whole time with physical explanation (in an excellent essay[12], Adrian Cussins calls this effect "Miraculous Coincidence"). More than enough should have been said by now however, to know what I would reply. The point of this analogy is really the same as the computer thing. If we share with the Sphex wasp, the feature of all our actions being explicable in mechanistic terms, then intentional explanation is as unnecessary for us, as it is for the Sphex wasp, and we can no more be held responsible for our actions than can the wasp. It may be a bit easier, due to the complexity of our mechanics to describe us in the intentional idiom, that's all.

As I've said, it's in "On Giving Libertarians What They Say They Want", that Dennett's efforts become a bit less mundane and more novel, if only because he tricks us all along by keeping what I think is the punch-line to the very end (although he does have grace

enough to be embarrassed by this). It appears initially that he is gainsaying a central thesis of "Mechanism and Reponsibility": this is because the task he sets himself in "On Giving Libertarians What They Say They Want" is that of finding a possible area of indeterminism which will at least help facilitate confidence in our belief that we are the authors of at least some of our actions. That is, he has claimed in "Mechanism and Reponsibility" that we are finite mechanisms, a clear consequence of this being that (whilst it may, in practice, be hopeless fantasy) all our thoughts and actions can, in principle, be accounted for wholly in the apparatus of physical causation, but in roaming around here for a useful area of indeterminism, his project appears to come into conflict with this mechanistic thesis. This, however, is where the trick is; as I hope to explain, what I think are odd, curious and ultimately futile liberties seem to be taken with concepts such as randomness, determinism and indeterminism and it should eventually become clear that there is, in fact, no real conflict between the respective positions taken up in the two essays. Mechanism (as should become clear) turns out not to be denied, after all, in the "libertarian" essay which I think (as I've already hinted) turns out to be little more than a slightly off-beat attempt at disparaging the ultimate importance of physical causation vis à vis agency and the related notions such as moral responsibility. Therefore, despite its title, the essay turns out not to be libertarian in the traditional sense in that Dennett ultimately recoils from denying physical determinism (indeed, he admits, after all, that he believes it still to be true) - it would be less misleading, I suppose, to call it compatibilist.

Dennett is well aware that, as well as the problem set for agency by determinism, there is also a problem set for it by indeterminism, i.e. if my actions are not determined by anything, then they appear to just happen at random, in which case there seems no reason either to hold me responsible for them, to regard them as mine. That is why Dennett's task is to find an area and kind of indeterminism which will help make it possible for me to see my actions as mine - not just any old indeterminism will do, since indeterminism could well just leave us with the old Libertarian's

Dilemma, which I've just elucidated. Indeed, I will consider later a very challenging compatibilist argument (the Mind argument) that free will in fact entails determinism, that the idea of a free act which is undetermined (or indeed, any undetermined act, for that matter) is incoherent. And in fact, before moving on to Dennett's explicit treatment of the problem, it's more than worthwhile, I think, to cite comments made by Van Inwagen in his preamble to the Mind argument, comments which clearly display the problem to which Dennett is addressing himself (the language strata claim, as I hope will become clear, goes fairly easily with all of the more formalised compatibilist arguments which I will deal with, such as Paradigm Case, Conditionalism and Mind). Van Inwagen asks us to suppose, for example, that there is exactly one undetermined particle of matter somewhere in the universe, and that it is far from any rational agent, the rest of the universe being governed entirely by strict deterministic laws. In that case determinism is, strictly speaking, false. But clearly, if determinism is incompatible with free will, so is the thesis that everything except one distant particle of matter is determined. Let us suppose, however, that at every moment many undetermined events take place inside each human body. But let us also suppose that these undetermined events play no role in shaping or influencing anyone's acts. If determinism is incompatible with free will, then so is this postulated state of affairs. It is crucial to be clear on the point (which is stressed by Van Inwagen) that if the question whether there are any undetermined events is relevant to the question whether we have free will, this can only be because the question whether there are undetermined events that shape or influence our acts is relevant to the question whether we have free will. Therefore the incompatibilist must believe that free will entails that there are undetermined events that shape our behaviour. And even then, not just any such event will do. Van Inwagen provides another example which clarifies this. Suppost that among the events that have shaped my behaviour are certain events that I have witnessed or experienced between the ages of one and twelve, events that were central to my moral education and to the formation of my character. Suppose that these events were somehow undetermined, but

that all subsequent events that made up my history were determined - or at least were determined given these earlier undetermined events. This too, incompatible with my having free will, provided determinism simpliciter is. Indeed, as is about to be borne out, Dennett is at least aware of the point Van Inwagen makes immediately afterwards, namely that there is only one point in the history of each of my acts at which the positing of any undetermined event could be conceptually relevant to the question whether that act was a free act: the point at which the act itself (or the deliberation that immediately preceded it) occured.

Dennett postulates an electronic "answer box", with two buttons, a Yes button and a No button, and two foot pedals, a Yes pedal and a No pedal. It also has a display screen divided in half and one side it says "Use the buttons" and on the other side it says "Use the pedals". Once a minute a radium randomiser determines, in an entirely undetermined way (this is how Dennett characterises it at this point anyway), whether the display screen says "use the buttons" or "use the pedals" (and this whole idea, again, of what appears on the display screen being supposedly "random", undetermined will be seen, I hope, once properly explicated, to be of monumental significance within free will/determinism, even if Dennett doesn't quite think so). The experiment: we draw up a list of ten very simple questions that have Yes or No answers, questions of the order of difficulty of "Do fish swim?" and "Is Texas bigger than Rhode Island?" We seat a subject at the answer box and announce that a handsome reward will be given to those who correctly follow all the experimental instructions, and a bonus will be given to those who answer all our questions correctly. Dennett poses the question of whether the physicist can in principle predict the subject's behaviour. We suppost that the subject is in fact a physically deterministic system, and further that the physicist has perfect knowledge of the subject's initial state, all the relevant deterministic laws, and all the interactions within the closed situation of the experimental set-up. Even the unpredictable behaviour of the answer box will infect the subject on a macroscopic scale with its own indeterminacy on at least ten occasions during

the period the physicist must predict. So the best the physicist can do is issue a multiple disjunctive or multiple conditional prediction. Dennett then asks us to consider whether the intentionalist can do any better and, by way of introducing what he calls his "cheap trick" tells us that yes, of course, he can (and I hope it will become clear that the argument of this essay is at the end of the day of the 'language strata' type I have been discrediting, with a few odd supplementations); given certain ordinary, straightforward, perfectly imaginable assumptions, the intentionalist can say something like: "The subject will give Yes answers to 1, 3, 5, 7, and 9, and the subject will answer the rest of the questions in the negative". And the reason, of course, that Dennett regards this as a cheap trick is that there is no real difference, as he points out, in the predictive power of the two predicitors, i.e. the intentionalist, for instance, is no more in a position to predict whether the subject will move finger or foot than the physicist is. Moreover, the physicist may well be able to provide a detailed conditional prediction of the skeletal motion of the subject. However, what Dennett regards as a crucial point (the precise import of this should become clear) is that what we are normally interested in, what we are normally interested in predicting, is not the skeletal motion of human beings, but their actions, and the intentionalist can predict the actions of the subject (at least in so far as most of us would take any interest in them) without (as Dennett puts it) "the elaborate rigmarole and calculations of the physicist". That is, whilst neither the physicist nor the intentionalist can tell us outright, unconditionally, exactly when the subject will move his finger, and exactly when he will move his foot, the intentionalist can, once he knows the questions and with the aid of some banal assumptions, tell us precisely what we're interested in, i.e. precisely when the subject will give a Yes answer and precisly then he will give a No answer. From the point of view of predicting an <u>action</u>, we are not interested in whether the subject uses the buttons or the pedals - all we are interested in is whether he answers Yes or No. The possibility of indeterminacy in the environment of the kind introduced here, and hence the possibility of indeterminacy in the

subject's reaction to that environment is something with regard to which the intentionalistic predictive power is quite neutral.

It is important, I think, to pause briefly at this point, and reflect on what exactly is going on here with Dennett. Again, we have, first and foremost, the distinction between the physical and the intentional modes of description. There is no obvious attempt on the part of Dennett to deny the predictive facility and insight of the physicist, there is no obvious attempt at denying that human beings are as physically deterministic as the next thing (remember, Dennett askes us to suppose that the subject "is in fact a physically deterministic system"), and so, in so far as these fundamentals are concerned, it seems no more clear than it has done in any other case where any sort of indeterminism which could allow agency to get a foothold, can get in. So far, then, the idea of human freedom seems to make no greater sense than I have already suggested it makes in the case of Davidson, Kim, Fodor, Honderich, Pears and the Dennett of "Mechanism and Responsibility". However, Dennett seems to think that he has an extra card up his sleeve here which he can play on behalf of the intentional and this comes really, in the shape of the supposed random, undetermined fashion in which the answer box instructs the subject to use either the buttons or the pedals. This, together with the power of the intentionalist to predict "what we are normally interested in" (i.e. actions) allows us, Dennett thinks, to insert some primacy of the intentional. What kind of primacy this is, however, and what consequences it has, if any, for human freedom, is by no means obvious at this point. As I've already said, I'll be wanting to argue that nothing, in fact, said by Dennett throughout this essay, on the supposed behalf of the intentional, is of any ultimate use for the concept of freedom. The 'language strata' thesis, given this form, will be seen (I hope) to be as misguided as ever, and in fact, I will pre-empt matters slightly by saying that what is really the first principle of Dennett's whole effort here, i.e. what finds its way into the analogy of the experiment as the "random" component within the answer box will be seen, in actual fact, to be a complete (and pointless) fraud. This can all be best elucidated, I think, by

going on now to document and consider the use which Dennett tries to make of his analogy.

The predictable enough extrapolation Dennett suggests is that we have something like the answer box inside the agent. We can begin by noting that it is a commonplace of action theory that virtually all human actions can be accomplished or realised in a wide variety of ways. It appears that Dennett is not a man for mundane and hackneyed examples, and he points our that there are, for instance, indefinitely many ways of insulting your neighbour, or even of asserting that snow is white (in precisely the same way that in our experiment, there was more than one way of answering Yes, and more than one way of answering No). And just as we were not much interested in whether our subject moved finger or foot by way of answering (all we were interested in was <u>what</u> he answered), we are often not much interested in exactly which particular physical motion accomplishes the act we intend. Dennett asks us to suppose that our nervous system is so constructed and designed that whenever in the implementation of an intention, our control system is faced with two or more options with regard to which we are non-partisan, a purely undetermined tie-breaking "choice" is made. Displaying again his predilection for example which will stick in one's mind, Dennett presents us with the picture of our standing in the supermarket, wanting a can of Campbell's Tomato Soup, all roughly equidistant from our hands. We are then asked to suppose that a perfectly random factor determines which can my hand reaches out for. Dennett then goes on to make remarks which will, again I think, be seen to be crucial to his enterprise, but lacking in the import for freedom which he thinks. First, he notes that this little area of indeterminism seems completely inconsequential. This is because it seems that it could only secure freedom for such a small and trivial class of our choices. What does it avail me if I am free to choose <u>this</u> can of soup, but not free to choose between buying it and stealing it? Dennett however, is convinced that he is onto something of importance here, though of course, I will be wanting to say eventually that he isnt't. We must not, he thinks, underestimate the possible scope of application of this idea; <u>for</u>

all we know (emphasis mine), which variation (in ways of implementing any action) occurs is undetermined, i.e. the implementation of any one of our intentional actions may encounter undetermined choice points in many places in the causal chain. And Dennett's next observation will again be seen to be of such significance for his efforts that I feel I ought to simply recount it:

> The resulting behaviour would not be distinguishable to our everyday eyes, or from the point of view of our everyday interests, from behaviour that was rigidly determined.

This should really, I think, provide something more than a whiff of what is to come, of what is actually going on. And in fact, at the end of this section, I will group together Dennett, Donald Davidson, J.L. Mackie and Keith Lehrer (who will be discussed in some detail later) as examples of a class of contributor within free will/determinism who display the same basic shortcoming, this shortcoming being pretty clearly displayed in the extract just quoted. As I hope to make much more explicit as I proceed, this general shortcoming has provided many a bogus reason for believing that compatibilism is true, a fact which Van Inwagen seems to appreciate. I feel that I cannot over-stress the importance of identifying precisely the kind of move which is being made by Dennett here, though I will nevertheless say that I think the best approach is simply to bear with him for the time being, and allow the picture to gradually emerge, i.e. the picture which reveals the four contributors I have cited (and Strawson is really another, at bottom) to be among those who commit the same compatibilist crime. We can detect, in this extract, a slightly undercover downgrading of the strict importance of physical determinism, a shift in the idea of what is supposedly primary, of what the issue is really all about. We can see that the shift is from strict physical determinism to observable distinctions which we can make; this strikes me, prima facie, as an odd shift in emphasis, precisely because we are perfectly well aware that the question of whether determinism is true, quite aside of whether it can be squared with

freedom, is not something which can be settled by banal observation of human behaviour. Indeed, the whole free will/determinism issue has such a grand philosophical history precisely because simple observation leaves the truth or falsity of determinism unsettled, and consequently (I think, as an incompatibilist) leaves the coherence of notions such as moral responsbility unsettled also. However, it seems quite clear here that Dennett seems to believe that the significant distinctions with regard to freedom, are actually to be made within simple observation, i.e. with no regard to the question of whether what is observed is actually wholly determined or not. And indeed I can now state what seems to me to be the punch-line of Dennett's story, which should come as little surprise now. As I've already said, Dennett, to his credit, regards it as something of an embarrassment, though as far as the aspirations of the libertarian are concerned, I think it is a great deal more than that. It concerns the alleged causal indeterminacy of the consideration generator. We have been supposing that the process generates considerations at least in part by a physically or causally undetermined or random process. However, throughout the essay, Dennett has made a lot of ambiguities in the word "random", and here he thinks that he can exploit another one. He notes that when a system designer or programmer relies on a "random" generation process, it is not (my emphasis) a physically undetermined process that is required, but simply a patternless process. Computers are typically equipped with a random number generator but the process that generates the sequence is a perfectly deterministic and determinate process. There will just be a complete absence of regularities on which to base predictions about unexamined portions of the sequence. And the "indeterminism" in the agent is really of a piece with this, i.e. strict physical determinism does not stop, but then, according to Dennett that is not ultimately important. According to Dennett, it is patternlessness which is important, and Dennett thinks that, as well as examples of patternlessness in the computer and "Campbell's Tomato Soup" kind of cases, there seems to be another case when we are faced with decisions to make, i.e. a consideration generator which is "random" in the required sense (patternless) is thrown into action. And as we shall see in a

moment, the question of whether this consideration generator is strictly physically deterministic is one Dennett holds to be of secondary importance - what is important for him is that it is patternless. It would have seemed reasonable then to ask Dennett whether he is giving libertarians what they want (or trying to give them what they want) without giving them indeterminism, but Dennett is clearly largely unperturbed by this prospect. As I've already said, he regards those who would view the ultimate presence of physical determinism as a calamity for libertarianism as mislocating the issue. Indeed, I think it worthwhile to quote outright a fairly large part of Dennett's reply to this misgiving, demonstrating clearly what Dennett takes to be the significant point of impact:

> Just as the presence or absence of macroscopic indeterminism in the implementation style of intentional actions turned out to be something essentially undetectable from the vantage point of our Lebenswelt, a feature with no significant repercussions in the "manifest image", to use Sellars' term, so the rival descriptions of the consideration generator, as random-but-causally-deterministic versus random-and-causally indeterministic, will have no clearly testable and contrary implications at the level of micro-neurophysiology, even if we succeed beyond our most optimistic fantasies in mapping deliberation processes onto neural activity.(emphasis mine)

There can be no doubt that something much more than trivial is going on here - specifically (as I've already suggested), there can be little doubt from this that Dennett holds the presnce of across the board causal determinism not to be by itself the killing blow to the head which the libertarian has traditionally taken it to be. Specifically, Dennett appears to by saying that beyond a certain level of complexity, the strict question of causal (physical) determinism, declines in importance, becomes secondary, as regards the libertarian's aspirations. It is at this point that Dennett is claiming there is no significant difference between the causally deterministic and the causally indeterministic - as we have already seen him suggest, the observations to which they would respectively

give rise would not be in the slightest bit different. So, according to Dennett, beyond the given level of physical complexity, whether the system is, in fact, causally determinstic or causally indeterministic, does not matter one iota to the libertarian, should not worry him in the slightest. And the significant point of complexity, the point beyond which the question of causal determinism or otherwise is inconsequential (for Dennett) is the point of patternlessness. Once patternlessness is observed, he thinks the libertarian can relax.

Whilst I have somewhat more specific questions to raise about Dennett's offering here, the central thrust of my remarks should be imaginable enough by now. That is, it, as the randomiser in the answer box was ultimately physically determined, the consideration generator in the agent is also physically determined, then I have to part company with Dennett right away and say that this fact is of itself of the utmost importance vis à vis freedom. The consequences or absence of consequences of this determinism for what we observe does not seem to me to matter; whether its upshot is something which is patternless seems to me inconsequential. What seems to me of the utmost importance is the mere fact of physical causation itself: if the consideration thrown up by the ("random") consideretaion generator "within" the agent are physically determined (à la the third premise of Honderich's determinism, say), then again it would seem to me to be very difficult to see what they have to do with the agent, how he is in any sense the author of them. Regardless of pattern or whatever, the intentional again seems to have its integrity supplanted by a more basic neurophysiological. That is why the supposed indeterminism, which could conceivably allow freedom to get a foothold, is ultimately a fraud. Dennett's consideration generator is, at the end of the day, as physically deterministic as anything, and, of course, I suppose it would have to be to square with the central thesis of "Mechanism and Responsibility", that humans are finitely rational mechanisms.

Indeed, Dennett seems to me to make rather odd and (as I've said) ultimately futile use of this notion of patternlessness-amidst-

strict-physical-determinism in order to try to restore the intentional to the position of primacy and integrity which the libertarian needs. Let me say first that this area of "indeterminism" seems to be the only one which Dennett seriously entertains: the consideration generator throws up considerations in a fashion which is physically determined but patternless and from there on in Dennett seems to think that the libertarian should be quite at peace with the idea of everything being wholly straightforwardly determined (I suppose this means "physically determined and not patternless"). As I've said, this really seems to me no different, when all is said and done, from a determinism so thorough-going as Honderich's and the consequent death knell for libertarianism. However, as I've also said, Dennett seems to do odd things with his notion of randomness, in order to try restoring the intentional to a position of primacy and giving flesh and blood to the concept of agency. And before actually saying anything about this manoeuvring on the part of Dennett, I will just repeat once more what I think cannot be over-stressed, namely that the random consideration generator <u>is</u> physically deterministic, and that between this essay and "Mechanism and Responsibility", there is every reason to think that Dennett holds everything which happens subsequent to the generation of considerations, to be physically determined. Dennett, remember, thinks that, despite the presence of what is across the board strict physical determinism, he has hit upon an "indeterminism", in the notion of the "undetermined" consideration generator, which is an aid to the libertarian cause. I have already said that I believe the presence of physical determinism to be a devastating blow to the libertarian. But even if it were not itself the case, I still find it difficult to see how the supposed "indeterminism" of the "random" consideration generator could help the libertarian one iota. On Dennett's model, even the most monumental decisions one will take will depend upon whatever considerations happen to be thrown up by a random generator, over which I am totally powerless. Why am I not, therefore, entirely impotent with regard to my actions? At any rate how <u>this</u> specific "indeterminism" posited by Dennett would help me to claim authorship of my actions is something over which I am in the dark. How would

it be any worse for the libertarian if the consideration generator was wholly unambiguously determined (I suppose this must mean, for Dennett's purposes "displays pattern"): how would this frustrate the notion of personal authorship, whilst the parallel "indeterminism" helps us, supposedly, to make a case for it? Consider for the moment, Dennett's illustrative example. Jones, who is finishing her dissertation on Aristotle and the practical syllogism, must decide within a week whether to accept the assistant professorship at the University of Chicago, or the assistant professorship at Swarthmore. We are asked to suppose that considerations A - F occur to her, and that those are the only considerations that occur to her, and that on the basis of those, she decides to accept the job at Swarthmore. She does this knowing of course, that she could devote more time and energy to this deliberation. Dennett also asks us to suppose that, after committing herself irrevocably, consideration G occurs to her, and that had G occured to her before she committed herself, she would have gone to Chicago instead. Dennett, strangely, seems almost manic in his desire to stress the intelligible nature of the action in the face of this possible "indeterminism" e.g. although we are supposing that the decision is strictly unpredictable, except conditionally by the intentionalist (despite its being ultimately physically determined), whatever choice Jones makes is retrospectively intelligible. There will be a rationale for the decision in either case; in the former case a rational argument in favour of Swarthmore based on A - F and in the latter case, a rational argument in favour of Chicago based on A - G. However, I don't myself see how this can possible elevate the intentional to the position of primacy and autonomy which the libertarian requires - indeed, it seems to me that this whole suggestion on the part of Dennett would be unpalatable to the libertarian. Let us recap. Considerations A - F have occured to Jones at the moment she decides where to go and so, on the basis of A - F, she decides to go to Swarthmore; however, had G occured to her before she made that fatal phone call, she would have gone to Chicago instead. This is surely one of the biggest decisions Jones will make in her life, and yet the eventual choice she makes will turn upon whatever

considerations (if Dennett is right), by processes over which she has no control, happen to get into her mind. That going to Chicago on the basis of A - G would have been no less and no more intelligible than going to Swarthmore on the basis of A - F is, I think, of little help to the libertarian. And Dennett, I think, simply compounds the implausibility of this idea by adding that there could be yet another rational argument, based on A - H, or I or J, in favour of Swarthmore, or in favour of going on welfare, or in favour of suicide. Not only may Jones choose to take the assistant professorship at Swarthmore on the grounds of considerations which spring up out of her control, but she may even decide to kill herself on the grounds of considerations which have a similarly autonomous origin! And this (Dennett thinks) raises the intentional to a position of integrity, helps us claim authorship in the face of physical determinism. This idea does not seem at all reasonable to me and indeed, the general suggestions made by Dennett about this example from which he thinks he can derive a useful "indeterminism" and primacy of the intentional, seem themselves to me to be difficult to swallow. It does not seem to me to be at all typical of important decision-making that a crucial consideration which carries such weight that it is sufficient reason to completely reverse one's original decision, occurs to one after one's decision has been made. Jones, Dennett suggests, considers the difference in salaries, the probable quality of the students, the quality of her colleagues, the teaching load, the location of the schools, and so forth. She then makes her decision (on the basis of A - F), knowing that she could devote more time and energy to this deliberation, could cast about for other deliberations, etc. And, as I've said, we're asked to imagine that after committing herself, consideration G occurs to Jones, which is enough to make her prefer the option she has already decided against. What seems to me to be wholly untypical about this is that, short of being given about two seconds in which to make up one's mind, all the major relevant factors occur to one more or less immediately, i.e. at first time of asking. The kind of considerations I've just documented Dennett as giving are precisely the kind on which a decision will be founded; these kind tend to occur en bloc, whilst the ones which don't occur straight

away will be relatively trivial e.g. the general quality of Indian restaurants in the area. The idea of something cataclysmic striking Jones after she has opted for Swarthmore seems to me implausible. I repeat however, that even if Dennett's remarks here are reasonable, how they provide a pathway for the integrity of the intentional and consequently for human freedom is not at all clear to me; indeed, as should be clear by now, the consequences of this are, I think, dire for freedom. Physical determinism seems just as haunting as ever.

Later remarks made by Dennett in supposed defence of the intentional make his efforts, I think, all the more perplexing. Near the end of his essay, Dennett says that

> in many cases our ultimate decision as to which way to act is less important phenomenologically as a contributor to our sense of free will than the prior decision affecting the deliberation process itself: the decision, for instance, not to consider any further, to terminate deliberation; or the decision to ignore certain lines of enquiry.

How is this feature connected with the "indeterminism" which Dennett has postulated as being instrumental for the libertarian? Let us suppose that Dennett is correct in his suggestion of the "random" consideration generator; the show gets on the road (let's say) with the random throwing up of considerations affecting the decision. On Dennett's model, it is perfectly possible however that everything that subsequently happens is quite straightforwardly physically determined (surely likely, in fact, given other general positions taken up by Dennett), and this must surely include the decision not to consider any further, to terminate deliberation, etc. And so, if these decisions affecting the deliberation process itself are themselves determined in the full-blooded sense (and of course, I would also say this even of the "weak" sense in which Dennett speaks of the random consideration generator), how can we be right in any feeling that they reveal the freedom of our will? Whether Dennett specifically thinks that the prior decisions affecting the

deliberation process are themselves determined, and what the relation of this to the random consideration generator is meant to be, is not clear to me, i.e. on the one hand, Dennett presents (as I've said, I think misguidedly) the random consideration generator as a condition of the success of the libertarian project and, on the other hand, he tells us that the decisions affecting the deliberation process itself are, more than anything, what makes us inclined to attempt the libertarian project in the first place. But of the connection between those two (alleged) aspects, he says absolutely nothing. I have already cast doubt on whether Dennett's suggested "random" consideration generator actually could reinforce any sense of personal authorship and, if I am correct about this, and also if everything which subsequently happens is determined, then the libertarian would seem to be back at square one i.e. the physical has won out, the intentional has lost, we have no freedom. And I think that similar remarks apply regarding another juncture which Dennett regards as having libertarian import, i.e. the intelligent selection from the considerations supposedly randomly thrown up by the generator - if, again, this intelligent selection is itself causally determined, then it seems to me that the game is up here also for the libertarian. Forget finding freedom here. But Dennett is conspicuously quiet on all this. One would expect him to provide an account of this "intelligent selection" which marks it as <u>mine</u>, in the same way as one may expect it of the (allegedly) phenomenologically important factor of the ending of deliberation, as one may expect on account also of how it is connected to what seems to be (for Dennett) the libertarian bedrock of everything, i.e. the "random" consideration generator. It is by no means clear how personal authorship is enhanced one iota by the supposed features Dennett takes to enhance it. Physical determinism, I think, still won't go away, and with it the consequences, which I think are every bit as monumental as we have seen Honderich say that they are. If physical determinism is true, if (in Ginet's terms) we have no choice as to which complete first-level description we are to satisfy, then despite the zealous and occasionally desperate attempts of 'language strata' theorists such as Davidson, Kim, Fodor, Pears and Dennett, we have no genuine choice at all regarding

our actions or thoughts - we cannot suddenly jump into this magical realm of the intentional to recover our freedom. Ginet and Van Inwagen seem right, despite everything. And in fact, this case against language strata compatibilism is reinforced succinctly in an extract of Malcolm's aforementioned essay "The Conceivability of Mechanism". Malcolm, by way of example, imagines the case of a man climbing a ladder to get onto a roof, in order to retrieve his hat, which the wind has put there. I have myself already said that any physical theory, if physical determinism is true (quite regardless of anything else), provides sufficient causal explanations of behaviour and Malcolm reinforces this, saying that a neurophysiological explanation of some behaviour that has occured is assumed to have the following form:

> Whenever an organism of structure S is in neurophysiological state Q it will emit movement M.
>
> Organism O of structure S was in neurophysiological state Q.
>
> Therefore, O emitted M.[13]

So as Malcolm goes on to point out, it would follow that the movements of the man on the ladder would be completely accounted for in terms of electrical, chemical and mechanical processes in his body. And, as Malcolm puts it:

> this would surely imply that his desire or intention to retrieve his hat had nothing to do with his movement up the ladder. It would imply that on this same occasion he would have moved up the ladder in exactly this way even if he had had no intention to retrieve his hat, or even no intention to climb the ladder. To mention his intention or purpose would be no explanation, nor even part of an explanation, of his movements on the ladder.

That is, it would not be true that those movements occured because he wanted or intended to get his hat; events in the remote past

conjoined with the laws of nature (neither of which are up to the man on the ladder), uniquely determine the movements of the man on the ladder, (or in Ginet's vocabulary, the man's movements are contingently necessitated through laws not of his making).

I hope that I have, by now, given some reasons for believing that despite the weight of philosophical writing against them, Malcolm, Van Inwagen and Ginet are correct regarding this consequence of the truth of strict, universal, physical determinism, with regard to human behaviour, though I will later be documenting further formalisations of Van Inwagen's incompatibilism and considering more challenging compatibilist arguments. 'Language strata' compatibilism, despite the toil and sweat which have clearly gone into articulating it, is ultimately indefensible. Later, I will consider (and ultimately reject) other compatibilist projects, specifically those which may, for shorthand purposes, can be called 'empirical' ones, those focusing on conditions of meaning and application of power-ascription statements e.g. that of Keith Lehrer in 'An Empirical Disproof of Determinism?' Other attempts at disparaging the supposed importance of the prospect of universal physical causation vis a vis our conception of ourselves as free agents will also eventually be considered e.g. that of Harry G. Frankfurt in "Freedom of the Will and the Concept of a Person"[14], where Frankfurt postulates the kind of person who "lacks nothing in the way of freedom", yet also says that the life of such a person is entirely independent of whether determinism is true. Before going on, however, I feel it worth considering a small contribution to the issue made by someone whose suggestions appear to display significant similarities with (in different facets) those of Dennett and Honderich. In the section of his well-known Ethics: Inventing Right and Wrong entitled "Determinism, Responsibility and Choice", J.L. Mackie (a philosopher whose works I tend to admire) argues for what is basically a compatibilist position, and a position which is similar to that of Honderich in that the determinism he concedes may be true is very similar to that of Honderich, but also similar to that of Dennett in that Mackie seems to believe the ultimate consequences of the truth of such a determinism to be fairly

minimal. Indeed, I think the comparison with Honderich is especially interesting. I will expound.

As the primary subject-matter is ethics, Mackie is, of course, concerned primarily with moral responsibility, and it is as one of the frontiers of ethics that the question of free will and determinism gets consideration. Mackie has less confidence than Honderich that the determinism he discusses is likely to be true - the argument he claims, is "far from watertight, but does provide the basis of a good case for determinism about actions". It should be clear that the argument consists of three premises very much like those of Honderich's ('One') determinism, namely that all physical states and events are causally determined, including states of the brain, that brain states are correlated with mental states in such a way that given a certain brain state just such a mental state must occur, and that actions are causally determined by mental states. Accepting the possibility of this being true, Mackie goes on to document the incompatibilist worry as the worry that choices are not open in any absolute sense. That is, they are fixed in advance because they are predictable in principle (despite the practical implausibility of such predicitions). The perfect predictor, as Mackie puts it, would be a relevantly exact replica of the agent and his environment, since determinism accepts the 'same cause, same effect' principle it implies that whatever choice the replica makes, the agent will also make (the echoes of Dennett's picture should be clear here). Mackie admits that this does go against our intuitive view of choice; we tend to assume, however unwarrantably, that decisions are not fixed in advance even in this minimal way. However, where I think Mackie is wrong here, and Honderich right, is that Mackie does not see why this should matter. Mackie wonders what it is that the free will apologist is worried about when he regards the consequences of this picture as damning: he asks whether the incoherent demand is being made that "the 'I' should be able to make that same 'I', itself at that moment, different from what it is". However, I don't think that this (incoherent) demand is the demand which is being made. The incompatibilist worry here is generated, not by the inability of the 'I' to make itself

different at any given moment, from what it is, but by the features of the world which make the 'I' precisely the 'I' which it is at that moment. That is, if the determinism documented by Honderich or Mackie is true, then the 'I' can, in principle, be exhaustively explained, at any given moment, by reference to physical states and physical laws (which are not up to me), the language of nerve cells, neuron firings, etc. How and when do I possibly get into the action i.e. if reference to my choices, desires, etc. are ultimately inessential to its explanation? Honderich is much more clear on this and with this in mind, I find it especially interesting to note Mackie's subsequent diagnosis of the incompatibilist worry:

> The most plausible interpretation is that it presupposes a distinction and contrast between my character, desires, and so on, all my contingent empirical features, and my real self - in Kantian terms between an empirical self and a noumenal or metaphysical self. The complaint is that if determinism holds, the empirical self is no doubt operative and effective, a cause as well as an effect, but the metaphysical self is an idle spectator of a causal order in which it cannot intervene.

I am as aware as Mackie is, of the difficulty of coherence created by the entire notion of the metaphysical self. And moreover, as I will make more explicit in a moment, so is Honderich. However, the interesting thing, I think, is that Mackie is mistaken to believe that there only is an incompatibilist worry if one attempts or wishes (or appears to fail) to find a role for a metaphysical self-type entity. As I've already said, Honderich is much less blase than Mackie about the possibility of encompassing physical explanation, he believes that it does present a genuine threat to our feelings about authorship and responsibility, but it should be clear from the following passage from "One Determinism" that this belief of Honderich does not itself turn upon a prior hankering for any kind of metaphysical self:

> It is not a matter of importance, but it is worth notice that to accept determinism is not to be deprived of a satisfactorily

articulated conception of responsibility. The blank occurs at that point where one should have an account of the non-causal agency that is ordinarily supposed to enter into responsible action. No one has ever offered more, by way of explanation, than a certain amount of dubious machinery, notably the 'Creative Self'. It is pretty hard to maintain the required suspension of disbelief in such items, or rather, it is hard to see what it is that one is trying not to disbelieve.

This is interesting, for a number of reasons. As I've suggested, Honderich clearly is as sceptical as Mackie as regards any metaphysical self. However, it seems that Honderich thinks that, if his determinism is true, then the kind of factors which make the empirical self the precise empirical self which it is at any given time (i.e. physical) require that we at least revise our conception of ourselves and our supposed responsibility for our actions. Mackie does not seem to grant this, since the truth of a Honderich-like determinism, according to him, needn't impinge one iota on our responsibility - ascriptions - one could, he thinks, only believe otherwise if one is already holding on to the mysterious notion of a metaphysical self. I think Mackie wrong (as I've suggested) to think the issue of the genesis of the empirical, contingent self irrelevant to questions of responsibility: if it is not up to me how my empirical, contingent self is constituted, then how can I make sense of my being responsible for my actions? I think it noteworthy that Honderich says only that his determinism would require some fundamental revision of ourselves (i.e. not a complete abandoning of all idea of responsibility), without actually spelling out what this revision would be like, how we could retain responsibility in the face of his ('One') determinism. Unlike Mackie however, he at least accepts that it would require some fundamental change in our conception of ourselves. Indeed, in two relatively straightforward, uncontroversial sections, "Voluntary or intentional actions" and "The straight rule of responsibility", which concern themselves more with our present practices and descriptions than anything else, Mackie believes that he makes out a powerful case for compatibilism. However, like a great many

compatibilist projects I have encountered (some of which I have already discussed, and some more of which I will discuss later), I don't believe that freedom or responsibility can be bought nearly so cheaply as the proponent believes. In Section 4, "Hard and soft determinism", Mackie notes that the distinctions drawn between intentional and non-intentional actions, the explanation suggested for the straight rule of responsibility, and the considerations that might justify real or apparent divergences from this straight rule, were all developed without even raising the question of contra-causal freedom. However, my reply to this is that what this does not show is that the question of contra-causal freedom is not relevant, that it does not impinge on what Mackie documents as having been fairly effortlessly discussed. That is, it may have been a mistake not to raise the question of contra-causal freedom and indeed, I believe (like Honderich) that if determinism is true, questions of responsibility must begin with an acknowledgement of that fact. Mackie seems to be guilty here of something which he warns against elsewhere, i.e. of believing that one set of facts do not impinge on the consideration of another set when, in fact they do. Indeed, anyone with general experience of the two of them may think it unlikely, but I think there is something rather Strawsonian in this approach of Mackie, i.e. he seems to believe that since we have an ethical life we want to get on with, that we do get on with, then questions regarding the rational justification of particular ethical stances can only make sense within the system itself, which is already given. It makes no sense to place the entire network in doubt via a question such as physical determinism. As I will argue later, I agree with Nagel that this view is incorrect. Indeed, before moving on, I think it will be very expedient for me to pause, do a bit of assimilation, and provide a foretaste of something which is to come a bit later, something which I hope will be seen to play a major role in supporting incompatibilism. When I come to devote some attention specifically to Keith Lehrer's variant of empirical compatibilism, I will cite a fanciful, but logically possible hypothesis postulated by Van Inwagen, a hypothesis Van Inwagen calls "(M)". All I will say about (M) at the moment is that it is a determinist hypothesis, the truth of which (it seems to Van Inwagen

and me, at least) would clearly preclude free will. Yet the point I want to make is that, having talked of Davidson, Dennett and Mackie, there is nothing said by any of them, the truth of which entails the falsity of (M). Everything said by Davidson, Dennett and Mackie regarding language strata, standpoints, practices, etc. could be true, and yet (M), also be true. One of the major objectives of Van Inwagen's in his floating of the (M) hypothesis, one of the points he is most anxious to make is one which I myself regard as of the most fundamental importance and will be stating repeatedly and in varying contexts throughout, namely that there are possible worlds in which things may appear exactly as they appear in the actual world, and yet one of the possible worlds in question be deterministic and the other one not. Identity of appearance or practice does <u>not</u> entail an identical answer to the question "Is determinism true?" This is a major problem, I think, with any kind of contributor who thinks that the question of determinism or compatibilism can be settled by an appeal to the world as it appears, to our language, practices and habits. Whether determinism is true is something we may only be able to settle by going beyond these primitive kind of features; hypothesis (M) is deliberately formulated so as to make explicit, not only that its truth would preclude free will, but also that its truth or falsity is perfectly compatible with the way the world appears to us, with the phenomena described by the likes of Davidson, Dennett and Mackie. Davidson, Dennett and Mackie could go on forever reiterating their points about language, standpoints, practices, etc. and yet not move one iota in the direction of establishing whether (M) is true or false. And in fact, I also think it of great importance to be clear that whilst (as Van Inwagen himself comments), there are significant differences in contrast between hypothesis (M) and the likes of Ginet's 'H' hypothesis or Honderich's ('One') determinism, the parallels are also crucial. Indeed, I think that a large part of the beauty of (M) for the incompatibilist is that, once one agrees that its truth precludes free will, then one would lose just about all reluctance to accept that the truth of H or Honderich's determinism would likewise kill off free will. Similarly, as Ginet and Honderich are no doubt perfectly well aware, one or other of

their determinist hypotheses could be true, or be found to be so, quite independently of how the world appears to us. As with (M), their respective suggestions can only really be discovered to be true or otherwise by probing beyond appearances and, as with (M) also, no amount of accuracy of documenting of language, standpoint or practice can counter-evidence them. Like (M), I think (as I will continue to adduce arguments for) the truth of either of them is imcompatible with free will, and like (M), I think that we cannot know their truth value, or counter-evidence their alleged incompatibility with free will though the descriptive analysis (however rigorous) of the likes of Davidson, Dennett, Mackie or (as I hope will become clear) Lehrer and Flew. For the moment, I think it very important to realise that, when presented with any argument for compatibilism or free will, we can always ask whether the argument invalidates Van Inwagen's First Formal (incompatibilist) argument, or whether it would help square free will with hypothesis (M), were it true. This is a highly useful approach, I think, and it seems to me that nothing offered by way of the contributors I have just been considering either invalidates Van Inwagen's First Formal Argument, or helps square free will with (M).

Indeed, I think that the significance of this kind of realisation really cannot be over-stressed, and believe that failure to grasp its importance has been at the bottom of a lot of what has been ultimately empty or insipid writing within free will/determinism. It should be clear that it has certainly been at the root of much compatibilist jinking. In fact, the approach of the likes of Davidson, Dennett and Mackie has an even more distinguished history than I have documented so far. In a well-known essay "Philosophy and the Scientific Image of Man",[15] Wilrid Sellars documents what is really a language strata/"standpoints" distinction, when he distinguishes between the <u>scientific</u> image and the <u>manifest</u> image of man. Not surprisingly, the former is one which we adopt when we wish to view ourselves as creatures in the natural world of cause and effect, and the latter is (supposedly) one we go in for when we wish to view ourselves as agents with intentions, purposes, objective, etc., and deserving of

praise, blame, gratitude, resentment and the like. The parallels with Dennett's mechanistic/intentional distinction in standpoints (or Strawson's objective/participant distinction in attitudes) are clear enough. As I've already indicated, I will later be looking at the comments of Nagel (with whom I am largely in agreement), on this kind of distinction and the use to which it has been put within discussions of free will/determinism. And, as well as Lehrer and Flew, I will also be considering another (to me) surprisingly influential and respected tradition within compatibilism which seems to me to harbour the same kinds of dismissive confusions, i.e. conditionalism. And, (to conclude the piece of preamble), let me say that, whilst Van Inwagen is highly useful in making this strand within compatibilism and its shortcomings explicit, an irony is that I will ultimately find Van Inwagen guilty of a crime not too different from the one which he has helped to describe, when I come to consider whether there is reason to believe that we have free will.

In the meantime, it may help us to establish some sort of attitude towards determinism if we could get some more tangible grasp of the kind of thing being talked about, some better grasp of the cash value of determinism.

By working his 'H' hypothesis through, Ginet allows us to see what determinism actually looks like in practice, how (if true) it would cash out in the actual world. Having established (I hope) that he is on a sure footing so far, we can now go back to Ginet, and see a thorough-going H, so to speak. Through seeing how this would look, we may be able to establish some confidence regarding whether it is reasonable or fanciful to think that it is true and more generally, whether it is reasonable or fanciful to think that anything like it could be true.

Notes

1. In "Freedom To Act" (Essays on Actions and Events).

2. In "Psychophysical Laws" reprinted in Lepore & McLaughlin, Actions & Events: Perspectives on the Philosophy of Donald Davidson.

3. In Essays on Actions and Events.

4. In Essays on Actions and Events.

5. In Synthese, 1974, Vol. 28.

6. In Essays on Actions and Events.

7. In Analysis, 42.1, January 1982.

8. In Essays on Freedom of Action, ed. Honderich.

9. In Brainstorms.

10. In Midwest Studies in Philosophy, 1981.

11. In "Freedom And Resentment".

12. In "Varieties of Psychologism", Synthese, January 1987.

13. In "The Conceivability of Mechanism".

14. Reprinted in Free Will, ed. Watson.

15. In Wilfrid Sellars, Science, Perception and Reality.

4. A WORKING DETERMINISM

I referred throughout the last part to Ginet's 'H' hypothesis as a straightforward physical determinism, and this may, I suppose, have been pre-empting matters very slightly. From what the complete first-level description 'B' contains, and what it does not contain, it would be a bit rash, I suppose, to conclude straight off that the 'A' slots could not, for instance, contain the likes of psychological descriptions. That H must be a straightforward physical determinism, however, is made admirably clear by Ginet himself. Ginet notes (as I have just done) that, as it stands at the moment, H contains no restrictions as to the kind of descriptions that may occupy the 'A' slots. However, he continues, were descriptions of certain sorts of psychological factors - desires, intentions, beliefs and the like, to be put in the 'A' slots, then stipulation (2) of H would not have a chance of being true. To remind ourselves, stipulation (2) is that

> A_1 contingently necessitates A_2, A_2 contingently necessitates A_3,...A_{n-1} contingently necessitates A_n, A_n contingently necessitates B, where 'B' is a complete first-level description of a temporal segment of a human being's behaviour, and 'A_1' 'A_2',...'A_n' are a series of antecedent circumstances.

Ginet suggests that the reason he makes this claim is that such psychological factors, though commonly used in explaining behaviour, can never be regarded as contingently necessitating that behaviour, because they must always fail to meet either the first or the third condition defining contingent necessitation. That is, if they are described in such a way as to leave it a logical possibility that a person should satisfy the description but behave in some way other than the one being explained (thus satisfying the first condition), then it will always be found that we have no reason to doubt that a person does, while satisfying the description, have a choice as to whether or not he shall behave in that way (thus failing to satisfy the third condition). And if, on the other hand, these psychological factors are described in such a way as to make us feel confident that no one satisfying the description has any choice, but

must behave in the way that they explain, then it will always be found that this is because the description entails that sort of behaviour (thus denying the first condition) - the connection will have been made necessary at the expense of its contingency. It is, I think, worth reflecting on this for a moment. I think that Ginet is correct here about psychological concepts and indeed, it seems to me that if one accepts the (Davidsonian) assumptions of indeterminacy of translation and the holism of the mental, then he cannot be other than correct. As Davidson has so famously pointed out, the intentional descriptions which are applicable to a subject derive their meaning, at least in part, from their position as a component part within a coherent system (network) of intentional characterisations, i.e. of purpose, beliefs, desires, goals, etc. In so far as an intentional ascription fails to cohere with other intentional ascriptions to the same subject, then we will be sceptical of it - something, at any rate, either the new characterisation or a previously held belief about the intentional network of the subject, would have to go. This holistic character of the intentional leaves us generally with a certain leeway and openness as to how to characterise a subject intentionally and explain his behaviour intentionally (indeterminacy of translation), and therefore there must always (as we have just seen Ginet point out) be a gap between, say, purpose and appropriate behaviour. Someone, for instance, may have a specific desire, but act so as not to fulfil it because of other held desires, or goals, or whatever.

Therefore, there is the gap between desire and appropriate behaviour, and our description of the subject as one with this particular desire, whilst correct, does not facilitate the conclusion that he will act so as to satisfy the desire. Conversely, we may correctly describe the subject with reference to the set of goals and purposes which we know to be incompatible with the satisfaction of this desire, and confidently predict that he will not attempt to satisfy it (if he does we will be sceptical about whether he actually has the broader set of goals and purposes). And I don't think that this suggestion is in conflict with what is said by Malcolm in "The Conceivability of Mechanism",

when he contrasts the neurophysiological and the intentional, i.e. that the relationship between purpose and action is not a contingent one, but is an a priori one. As I considered to some degree earlier, the relationship between purpose and behaviour can, I think, be a priori, conceptual, without compromising the claim we have just seen Ginet make, or any of the Davidson premises. This, in fact, receives some echo in Malcolm's own example (which I have already alluded to) of the man trying to retrieve his hat from the roof. Let us look, for a moment, at the form Malcolm suggests an intentional explanation of the man's climbing the ladder would take:

> If a man wants to retrieve his hat, and believes this requires him to climb a ladder, he will do so provided there are no countervailing factors. (emphasis mine)

This man wanted to retrieve his hat and believed that this required him to climb a ladder, and there were no countervailing factors.

Therefore he climbed the ladder.

The notion of 'countervailing factors' hints strongly at the Davidson premises, and at what Ginet has said. And indeed, the echo becomes even stronger when Malcolm goes on to ask what sorts of things might be included under 'countervailing factors' in such a case. His suggestions include the unavailability of a ladder, the fear of climbing one and the belief that someone would remove the ladder while he was on the roof, i.e. more components within the network of the man's intentional states. I hope that I have helped make it clear here why Ginet is correct to say that psychological descriptions are precluded from the 'A' slots and why this does not mean disclaiming the a priori relationship between intentional states and action.

Indeed, perhaps I can really make it clearest of all by saying that all that Ginet is really saying with regard to the relationship between psychological states and action, what it really amounts to, is that (what Kim has called) Davidson's thesis of Psychophysical

Anomalism is true. (What Ginet is saying may or may not also entail that Psychological Anomalism is true, but I don't think it essential to pursue this presently.) For reasons which have already been discussed, the psychological is just not the kind of thing to enter into lawlike relations with the physical; because of features of the mental such as rationality maximisation, indeterminacy of translation, and holism, psychological concepts are nomologically incompatible with behaviour descriptions. If we go back a little bit, what Ginet is actually pointing out is that which recurs throughout Davidson's work, namely that any appearance of genuine lawlike relations between the psychological and behaviour must be illusory. Like Davidson, however, Ginet is perfectly well aware that there can <u>be</u> such an appearance and in the face of Psychophysical Anomalism, what he does is account for this appearance, and explain why it must be illusory. This is, really, the import of his comment that psychological factors can never be regarded as contingently necessitating behaviour, because they must always fail to meet either the first or the third condition defining contingent necessitation. The illusion of lawlikeness, in any given case, can be shattered by exposing the fact that the "law" in question has only been purchased at the price of having the psychological description <u>entail</u> the type of behaviour in question. And, of course, this has to be the case, given Psychophysical Anomalism i.e. any resistance, any apparent counter-evidence to a purported psychophysical law must be "explained away", and explained away again, ad infinitum, until the psychological description entails the behaviour in question, and the "law" collapses into nothing. It can, I hope, be seen from everything I have already discussed, especially vis a vis Davidson and Ginet that these are two mutually supporting features, they are not accidentally placed together, i.e. since we see, on inspection, that any psychophysical "law" is explained away into nothing, then we get very good evidence (on those grounds) for Psychophysical Anomalism, and (proceeding in the other direction) since Psychophysical Anomalism is something which seems to be true (on other grounds), then it is to be expected that any supposed psychophysical "law" can only be sustained by explaining it away into a stipulation. Psychophysical "laws" must

be stipulations because Psychophysical Anomalism is true, and Psychophysical Anomalism seems to be true because we find that an attempt at discovering a psychophysical law collapses into our fixing them by stipulation. What is really significant here is the result of things which are going on at the ground-floor; there are, really, ground-floor reasons why the 'A' slots within Ginet's 'H' hypothesis cannot contain psychological descriptions. Indeed, these ground-floor features (or equivalent ones at any rate) will become important again later on, when I come to address what I have called "Empirical Determinism/Incompatibilism". Within this species of determinism/incompatibilism, the "explaining away" manoeuvre is probably easier to identify than it is in the kind of case I would imagine Ginet (or Davidson) have specifically in mind (indeed, I will first discuss "explaining away" arguments in general), but the reasons why empirical determinism/incompatibilism can only be supported (I will conclude) by "explaining away" means are largely equivalent, i.e. that the kind of alleged relations, put forward to support it contain the wrong kind of thing to get into <u>laws</u>. And, as has been reinforced strongly enough up till now, and will be reinforced more as I go on, if there is one thing which determinism <u>does</u> entail, it is laws.

It now seems little more than trite to say that the sets of circumstances that might be candidates for the 'A' slots in H are restricted to the <u>physical</u> sort, and Ginet's follow-through, as I've suggested, shows precisely what the consequences of physical determinism would be, for how we would be forced to think of ordinary human conduct. Ginet states that we could imagine future observations that would make it reasonable to believe H, or to <u>accept</u> it, in the sense in which it is commonly said of an hypothesis in a science that it is accepted. He asks us to consider the following first-level descriptions of common behaviour sequences:

> The body moves from a standing position to position of being seated on a chair, and then the left leg rises and crosses over and settles on the right leg.

> The head turns slightly from left to right...and at the same time there issues from the mouth the sounds "In this paper I shall be concerned with..."
>
> The hand and arm move towards a pen...and then the hand and fingers move in such a way that the pen traces on the paper the following marks: "My Last Will and Testament..."

Ginet then goes on to ask a whole series of rhetorical questions about this behaviour sequence he has just imagined:

> Isn't it entirely conceivable that for each of these behaviour patterns we might observe that a certain sort of internal state of a human being (one, say, that is physiologically and neurologically very, very complex), when occuring in certain environmental circumstances, is always accompanied by that sort of behaviour, no matter how we vary the other circumstances? And isn't it just as conceivable that we should discover for each such physiological-cum-environmental state (that appears to necessitate contingently a certain sort of behaviour) a kind of antecedent set of circumstances that <u>it</u> always and inescapably accompanies? And so on, until we arrive at a set of antecedent circumstances that contingently necessitates a considerable chain of bodily-cum-environmental changes, ending in the sort of behaviour sequence we started with, and with regard to each of which remote antecedents it is obvious that the person whose behaviour is at the end of the chain could have had no choice?

It is, I think, worth repeating that, whilst this may appear particularly rough and ready to some and, I would expect, bizarre to others, it is my contention that, despite all their desperate squirming and writhing about, the anomalous monists, Dennett and Pears (Honderich, too) are committed to something much like this suggestion of Ginet. Despite the zealous efforts of those contributors to preserve some sort of integrity for the intentional and consequently for freedom, they must, ultimately, seriously

entertain something like the suggestion I have just cited Ginet as making. Ginet stresses that this is a logical possibility. And (he continues), we can imagine accumulating such observations for a great many different kinds of behaviour-sequences, until the grounds become strong enough to justify accepting H as a well-confirmed hypothesis.

Can this picture put forward by Ginet be swallowed, ultimately? The first point which I think worth making is not itself an argument but it is really only a persuasive point - I feel, however, that it is germane to what looks like the real stumbling block to an acceptance of H. This is simply that it does seem prima facie fanciful to imagine observations taking place which would make it reasonable to accept H. Perhaps my own imagination is just not so dynamic as Ginet's, but the suggestion that we could even confirm a relationship of contingent necessitation between one set of purely physically described antecedent circumstances, and one complete first-level behaviour-kind, doesn't seem especially hopeful to me, so the prospect of discovering enough such relationships to accept H seems to me completely wild. Take his own example: is it really that easy to imagine that we might arrive at the discovery of (say) some extremely complex internal state of a human being which contingently necessitates the cited movement of the body, the left leg crossing over the right leg, etc., which in turn contingently necessitates the sounds "In this paper I shall be concerned with...", and all of which, through a chain of contingent necessitation, ultimately contingently necessitates the pen tracing on the paper the marks "My Last Will and Testament"? As I've said, perhaps the reason why this seems to me so difficult to imagine hints at the real problem for something such as the H hypothesis, and that is precisely the essential presence of the intentional to an exhaustive account of the world. And I don't think I can over-stress "essential" here. As I've already said in connection with a number of manoeuvres on behalf of free will, it is of no consequence whatever if the intentional manages to get into the picture in some fortuitous fashion which leaves it ultimately secondary and inessential, i.e. if the behaviour which is supposedly

being explained would happen anyway, given the purely physical/ neurophysiological conditions holding at the time, leaving the behaviour in question wholly explicable in the absence of intentional characterisation of the subject. This would be the case, I said, if the pictures suggested by the likes of Davidson, Kim, Fodor, Pears, Dennett and Honderich were correct - everything which goes on in the world could be explained without any reference whatever to the intentional mode. As we've already seen Malcolm say in "The Conceivability of Mechanism", if this were the case, then it would not be true that the movements of the man in his example occurred because he wanted or intended to get his hat. It should be clear by now, of course, that Ginet's H hypothesis has this upshot, and it is precisely this feature which makes it seem impossible to accept. In Malcolm's example, it seems simply bizarre to suggest the man's movements did not occur because he wanted or intended to get his hat; it seems that the man's desire or intention to get his hat, as described in precisely this way, is indispensable in explanation of his movements. And, of course, equivalent remarks seem to apply in the case of Ginet's example - it seems (prima facie at least) that the series of antecedent circumstances relevant to an explanation of the whole affair must include intentional concepts. The subject in this example is (presumably) giving a paper to an audience. The circumstances which, it seems, must be cited by way of explanation of this include the likes of his being asked to give a paper by someone in charge who thought that his audience would wish to hear and would enjoy such a paper, the person's understanding of this offer, his subsequent acceptance of it, research into the topic, further work, planning, pondering, revising, etc. This kind of facet would seem to be an essential part of the person's body being where it is in the first place. And we could go on: the person sits on a chair perhaps because he wants to, or because he feels slightly fatigued after having to run to catch his train (this itself being the result of his becoming too absorbed in a book in the waiting room), because his audience would prefer it or whatever. His head and eyes move as they do because he wants to read his paper to his audience, his mouth only emits the first sounds that they do because the sounds have a meaning, which

both he and his audience understand, and because of accepted conventions of paper-giving; the hand moves towards the pen because he knows that the pen is an instrument with which to write, and he wants to write a set of meaningful marks on his paper. And so on. As I've said, it seems that those kind of features, whose essential nature seems to be as they are described (i.e. in the intentional mode) cannot be left out of a complete account of what happens. But if H is true, then they can be left out. This makes H look unlikely.

Purposive states qua purposive states seem to be essential to an explanation of action and so, if my arguments so far have been sound, determinism cannot be true. However, a rather unlikely and particularly radical source of danger to the entire schemata assumed so far could, if defensible, throw everything into some disarray. I can hint at it first by citing a rather lengthy extract from Book V of Nietzsche's The Gay Science:

> Even the most circumspect...think that the known is at least more easily understood than the strange...Error of errors! The known is the accustomed, and the accustomed is the most difficult of all to "understand", that is to say, to perceive as a problem, to perceive as strange, distant, "outside of us"...The great certainty of the natural sciences in comparison with psychology and the criticism of the elements of consciousness - unnatural sciences as one might almost be entitled to call them - rests precisely on the fact that they take what is strange as their object: while it is almost like something contradictory and absurd to wish to take generally what is not strange as an object... .[1]

Nietzsche was making a point specifically about the Cartesian project here, but it is this general kind of lead Nietzsche gives which is behind the endeavour I now wish to consider. All along, it has been tacitly assumed that we have straightforwardly intentional states, a sub-class of which are purposive states, a central question for free will being whether purposive states actually have any genuine application to the world, or whether we could explain

everything that happens without them. However, is this assumption itself well-grounded? What do we mean when we say that we have purposive states, and what do we know about them? If the whole idea of purposive states can be discredited, then perhaps a crucial part of the apparatus which generates the free will problem will have gone, and we will have to start thinking of ourselves in wholly new terms and, who knows, perhaps even terms which will transform the whole traditional problem of free will for us. For the eliminative materialist Paul Churchland, such a move seems to be a possibility, indeed, in his paper, "Eliminative Materialism and the Propositional Attitudes", Churchland puts forward the idea that "intentional states" are, quite conceivably, an eliminable construct of what he calls "folk psychology", and could well be replaced by something much more useful. As we shall see, there are two essential components in his disparaging of intentional states:

(i) that folk psychology, of which intentional states are a part, is a <u>theory</u>, and

(ii) that, given what is required of a theory, folk psychology is a barely credible theory.

It should be clear that, if these claims can be sustained, then the free will problem may take on a whole new complexion, or may even disappear. I will be forced, ultimately, to disagree with Churchland but, nevertheless, his efforts have to be taken seriously, and I will now turn to them.

Notes

1. In 355: "The Origin of our Conception of 'Knowledge'".

5. ELIMINATIVE MATERIALISM AND THE PERSISTENCE OF PURPOSIVE STATES

In order to properly place the threat of eliminative materialism in perspective, it is, I think, worthwhile to recap on features of the picture which I have already described. Away at the beginning, I talked of the primitive, preverbal kind of intuitions which are at the bottom of a great many of the classical philosophical problems, of which free will is one. I attempted to clarify this notion with regard to freedom, saying that what really generates a problem about freedom is the persistence of two conflicting impressions, i.e. on the one hand, the apparently unavoidable impression, on a great many occasions when we act, that we are the authors of our actions, that it is within our power to do other than what we, in fact, do, and that those actions (with regard to which are free) cannot be exhaustively explained without recourse to the irreducible intentional categories, and on the other hand, the almost equally unavoidable impression that, being (at least) physical objects in a natural world, cause and effect and our mental lives being, in some way at least, intimately related to the functionings of a brute physical entity (i.e. the brain), on each occasion on which we act, we are merely the focus of causal summation for external influence, leaving it very difficult to attach any sense to the idea that we could have done otherwise and that we are responsible for any of our actions. The feeling that it is the latter intuition, and not the former one, which points the way to where the truth lies, finds formalisation in the likes of Ginet's H hypothesis, which explicitly states the intentional to be (ultimately) inessential to explanation of any human action, and Honderich's ('One') determinism, which may not, prima facie, appear to be so hard-line regarding the intentional as Ginet's H, but which really has exactly the same significant consequence, i.e. the explanatory superfluity of the intentional.

I went on to say that it was precisely this failure on the part of a hypothesis such as Ginet's H to appear to do any sort of justice to the former intuition above, as opposed to the latter one, which made it look so implausible. This feeling was reinforced by

looking at a rigorous working-through of H, and in particular, at Ginet's purported illustrative example of H in practice; I said that it seemed essential to an explanation of the particular goings-on documented by Ginet that (for instance) the speaker understood the invitation to speak, knew something about his topic, etc., and an entire set of other things similarly constituted essentially by irreducible intentional concepts. The idea that these intentional categories could, in principle, disappear in favour of something more basic (i.e. the neurophysiological) seemed implausible in the extreme. Now enter what must be regarded as a knight in shining armour for someone impressed by H, i.e. Eliminative Materialism. I have already made clear, I hope, that what appears as a central strand in the worry about freedom is the fear that, whilst the intentional may have unambiguous, straightforward, full-blooded reality (e.g. I really do have an intention to go home), it is nevertheless a function of something yet more basic, i.e. the neurophysiological, which could, in principle, be invoked to explain exhaustively. I hope that I have also made clear by now that it is precisely this seeming full-blooded reality of the intentional, and its apparently essential links with behaviour which we have already seen Malcolm speak of, which makes this notion of a more basic neurophysiological seem unlikely. However, it is the business of eliminative materialism to get a foothold even prior to all of this, i.e. not to wonder (as Honderich clearly would) whether the neurophysiological may be more basic than a patently real intentional, but to facilitate the explanatory comprehensiveness of the neurophysiological by casting doubt on the entire presupposition of the intentional itself. As Paul Churchland puts it in the opening sentence of "Eliminative Materialism and the Propositional Attitudes":

> Eliminative Materialism is the thesis that our common-sense conception of psychological phenomena constitutes a radically false <u>theory</u>, a theory so fundamentally defective that both the principles and the ontology of that theory will eventually be displaced, rather than smoothly reduced, by completed neuroscience. (emphasis mine)

My emphasis will be seen, I hope, not to be gratuitous. One prong of my attack on eliminative materialism will be my claim, which I regard as of the utmost significance, that what Churchland calls "our common-sense conception of psychological phenomena" is not, in fact, a theory in any serious sense (Churchland calls this alleged theory "folk psychology", or FP). The other prong of my attack will concern Churchland's specific treatment of what, really, I have documented as one of the apparent intuitions giving rise to the whole issue of free will, i.e. propositional attitudes (beliefs, desires, intentions, purposes, etc.). As I've already said, the entire assumed existence of the likes of desires, intentions and purposes is something which Churchland thinks is, at best, presumptious, and this claim is really part of his more general claim that FP is a theory, and one whose credentials are, at that, in serious doubt.

Churchland spends a few pages attempting to support his claim that it is in fact only a theory that humans have desires, intentions, beliefs, purposes, etc., features which are (as Malcolm stresses) conceptually linked to their actions. However, I don't myself see that Churchland says anything in serious support of his claim that what he calls "folk psychology" is only a theory. I don't have the space here to devote to a consideration of the essential nature of a theory. However, I don't think either that this is quite essential. All we need know, I think, are the rough and ready features of FP, and whether these features problematise intentions, desires, beliefs, etc., in the way in which Churchland supposes; that is really, I think, what is essential to Churchland's claim that FP is a theory, i.e. that intentions, desires, beliefs, purposes and the like are not (as seems to be generally supposed) unambiguously real, "untouchable", <u>given</u> features of reality, but that such notions only derive any import through their position in an entire holism of concepts and relations (Churchland's "folk psychology"), which may itself be radically unsatisfactory. As I've said, Churchland cites some features of the mental, features which he thinks give the notion of intentional states this questionable credibility. He characterises FP generally:

> Each of us understands others, as well as we do, because we share a tacit command of an integrated body of lore concerning the lawlike relations holding among external circumstances, internal states, and overt behaviour.

In general support, Churchland notes the the average person is able to explain, and even predict, the behaviour of other persons with a facility and success that is remarkable. Such explanations and predictions standardly make reference to the desires, beliefs, fears, intentions, perceptions, and so forth, to which the agents are presumed subject. But such explanations (Churchland goes on) presuppose laws - rough and ready ones at least - that connect the explanatory conditions with the behaviour explained. Similarly, Churchland thinks that the intentionality itself of mental states has a clearly theoretical status:

> ...the intentionality of mental states here emerges not as a mystery of nature, but as a structural feature of the concepts of folk psychology...
>
> e.g. (x) (p) (q) [((x desires that p) & (x believes that {if q, then p}) & (x is able to bring it about that q)) > (barring conflicting desires or preferred strategies, x brings it about that q)]
>
> (Churchland's example.)

However, it is my own view that, far from the kind of features I have just cited Churchland as documenting, pointing to dubiously theory-laden conclusions we (allegedly) leap to regarding the mental and its connections with external circumstances and behaviour, these features are, on the contrary, facts about the <u>mental itself</u>, about the mental qua mental. Intentional states are not a theoretical postulate, but are a part of the full-blooded reality of mental phenomena; similarly, the kind of "rough and ready" laws which Churchland speaks of relating the intentional with behaviour are not merely components of an entire theoretical holism which we may be able to dispense with, but are a part of the essential nature

of actual real intentional states e.g. (in Malcolm's already cited example), it may be a part of a man's mental life at a given time that he wants to retrieve his hat (i.e. nothing "theoretical" here), that he (similarly) believes this requires him to climb a ladder and, if this is true, he will climb the ladder provided there are no countervailing factors (as an essential part of the nature of intentional states). (An excellent and thorough discussion of this, which I don't feel it necessary to pursue here, occurs in John Searle's Book Intentionality, Chapter 1, "The Nature of Intentional States" and Chapter 3, "Intention and action"). In his aforementioned essay "Analytic Philosophy and Mental Phenomena", John Searle makes, I think, what are very persausive remarks on this issue. He concerns himself at one point with the functionalism of Dennett, Dennett's position being one with significant similarities to that of Churchland, in that the actual, real existence of intentional states as an essential feature of the mental seems cast in doubt and instead, we are left only with "intentional systems" and the "intentional stance". On Dennett's account both humans and computers with certain sorts of program are "intentional systems", and an intentional system is just one where we find it appropriate to adopt "the intentional stance". In the adoption of the intentional stance, "One predicts behaviour...by ascribing to the system the possession of certain information, and supposing it to be directed by certain goals and then working out the most reasonable or appropriate action on the basis of these ascriptions and suppositions" ('Toward a Cognitive Theory of Consciousness'[1]). It is then, Dennett says, but a small step to describe this information and these goals as beliefs and desires. But we must not ask if these really are beliefs and desires because "the definition I have given of intentional systems does not say that intentional systems really have beliefs and desires, but that one can explain and predict their behaviour by ascribing beliefs and desires to them, and the decision to adopt this strategy is pragmatic and is not intrinsically right or wrong". The echoes with Churchland's suggestion that FP is a theory should be unmistakeable. I am, however, in agreement with Searle, who agrees that it is open to anybody to adopt any strategy he finds useful, but adds that the question remains (my emphasis),

what is the status of the ascription of the mental states, whether the ascription is of information and goals, or beliefs and desires. Indeed, I hope it is clear by now why I regard it reasonable to simply quote outright the following crucial sentence of Searle's:

> Even if we have not defined intentional systems in such a way that we say that they really have beliefs and desires, there will still be a difference between those intentional systems that really have beliefs and desires and those that do not, and in the case of those that do, the ascriptions have entirely different interpretations from the case of those that do not. (emphasis mine)

It is this definition drawn above by Searle which Churchland (along with Dennett) is questioning. According to Churchland's thesis that FP is a theory, there is no significant difference whatever between our ascriptions of intentional states to ourselves and to the computer - in both cases we simply employ a holism of concepts which, as it happens, has some efficacy in explaining respective behaviour. As I've said however, the likes of Churchland and Dennett appear to me to be completely in error here. As Searle comments, there is an enormous difference between my attributing a "desire" to a chess-playing computer to castle on the strong side, and my saying I have a desire to drink a glass of cold beer:

> ...in my own case I do not attribute to myself a desire for a beer because I find it useful in predicting my behaviour, but because I want a beer (emphasis mine)...whether or not people find it useful to adopt "the intentional stance" towards me is quite irrelevant to what the facts really are. (emphasis mine)

Indeed, in an ingenious analogy a bit later, Searle, I think, exposes the absurdity of the claims of Churchland and Dennett, that the alleged existence of mental phenomena such as beliefs and desires with intrinsic relations to behaviour is mere theoretical, not (in principle) indispensable holism. Suppose that there were a group of philosophers who were puzzled by the existence of hands.

And suppose that we were told that we do not have to worry about the existence of hands because it is all a matter of adopting the "manual stance" toward certain systems which we will describe as "manual systems". We can paraphrase Dennett here: "the definition of a manual system does not say that manual systems really have hands but that one can explain and predict their behaviour by attributing hands to them, and the decision to adopt this strategy is pragmatic and not intrinsically right or wrong". It should be perfectly clear by now what the point is here, namely that the intentional stance approach to understanding beliefs and desires is about as useful as the manual stance approach would be to understanding hands. As Searle points out, in each case the question of analysing the intrinsic features of mental states (or hands) gets replaced by a different question: under what conditions do we find it useful to talk as if a system had mental states (or hands)? As I've said, this is the distinction which the likes of Churchland and Dennett are doubting, but Searle, I think, reinforces it even more strongly by the penetrating observation that we understand the metaphorical non-literal use of "belief" and "desire" as applied to computers because we see these attributions as based on an analogy with systems such as human beings who literally have intrinsic beliefs and desires. That is, despite the desperate efforts of Dennett and Churchland, I think that Searle is clearly correct here: there is nothing "metaphorical" or "theoretical" in the ascription of mental states to human beings who simply have unambiguously genuine mental states, some of which have as an essential feature (as we have seen Malcolm say), a relation to action. The ascription to the likes of computers, on the other hand, of mental states, is metaphorical (or "theoretical"), in that it is merely a part of a convenient holism of concepts used to explain and predict behaviour, and is itself (as we have seen Searle point out) parasitic on the primary, non-theoretical ascription of intentional states to human beings. And whilst it is not, strictly speaking, an argument, I find a more than persuasive afterthought Searle's suggestion that one thinks what it is actually like to have a pain in the stomach or a passionate desire for a cold beer and ask one's self whether one is talking about real mental phenomena or

just adopting a stance. That is, we must ask what is the source of the explanatory categories which constitute the 'intentional stance'. According to the likes of Dennett and Churchland, the intentional stance would appear to be something which appears, as though by magic, from nowhere in particular, and which, somehow or other, allows us to predict the behaviour of 'intentional systems' i.e. as well as the physical categories, there appears to just spring up the entire intentional network of beliefs, desires, purposes, thoughts, etc. which we somehow are warranted in applying, in exactly the same way to some human beings, some animals and some machines, in order to predict and explain their behaviour. We just impose it (it goes on) upon certain classes of things and lo and behold, we find that it works sometimes in order to explain their behaviour. This, however, seems to me (and Searle) to be wholly unacceptable - it seems to me ludicrous to assume that, when in some cases we find the physical categories too cumbersome to explain and predict behaviour, we can simply hire, as it were, this other, completely different set of categories (i.e. with some humans, some animals, and some machines) for the purpose, without ascribing any sort of primacy to one kind of application over another, and leaving us completely in the dark as to where this other and curiously useful set of categories has come from in the first place. On the contrary, I would claim (as Searle would appear to), that we know perfectly well where the intentional categories, and all their logic, come from, and that is from the case of human beings, who literally have intentional states. This is where the primary use of the intentional categories is to be found, and the ascription of intentional states to machines is a metaphorical ascription, made possible because we already have, prior to this, the network of intentional concepts, which is a result of our literally having intentional states. Indeed, this whole claim on the part of eliminativists, that no particular primacy can be given to any specific application of the intentional categories, seems to generate the following infinite regress problem: if it is only under some interpretation or other that a physical set-up can be credited with mental states, is it only under some interpretation that what is offered is an interpretation? That is, it is suggested that it is

only under an interpretation (apparently arbitrary, as I've said) that even human beings can be ascribed mental states (i.e. the adoption of the intentional stance); however, if thoughts, beliefs and the like can only be ascribed under a particular interpretation, then an interpretation, being of the same category in this respect as thoughts, beliefs, etc. (intentional) can only be ascribed under a particular interpretation (the intentional stance). That is, it is claimed, on the one hand, that the intentional stance is only an interpretation, but on the other hand the intentional stance seems to be a condition of recognising something as an interpretation. The intentional stance, according to this, seems to be both a condition ascribing the character of an interpretation to something, but also an interpretation itself, which seems incoherent. Indeed, this seeming primacy of the intentional, this basicness, will eventually turn out to be of great importance, with regard to what conclusions I will ultimately reach regarding free will. In the meantime, we can get clear on the confusions of Dennett and Churchland by insisting at least at some point on a first-person point of view.

If my argument has been sound, then what makes Ginet's 'H' hypothesis seem implausible (and, of course, several variants of determinism, which despite disguise, I have claimed are significantly similar), namely the seemingly essential role of intentional concepts in explanation of human action, cannot be explained away via the claim that this feature itself is merely a component of an overarching theory which may well be defective and replaceable. As I hope I have successfully argued, what generates the implausibility of a theory such as 'H' is not a theory which the truth of H could itself falsify, or which may be false on independent grounds anyway, but unambiguously real features of the world. Indeed, if we have a look now at some of the points Churchland advances in favour of his claim, not only that FP is a theory, but that it is a radically defective theory, what I hope will be reinforced are precisely the counter-claims for which I have been arguing.

The first point Churchland attempts to advance to the supposed discredit of FP is the alleged failures of FP. He cites a whole set of apparently large-scale shortcomings: the nature and dynamics of mental illness, the faculty of creative imagination, the ground of intelligence differences between individuals, the nature and psychological functions of sleep, the common ability to catch an outfield fly ball on the run or hit a moving car with a snowball, the internal construction of a 3-D visual image and the subtle differences with the 2-D array of stimulations in our respective retinas, the rich variety of perceptual illusions (visual and otherwise), the miracle of memory, with its lightning capacity for relevant retrieval. On these and many other mental phenomena, FP, claims Churchland, sheds negligible light. He adds that "one particularly outstanding mystery is the nature of the learning process itself, especially where it involves large-scale conceptual change, and especially as it appears in its pre-linguistic or entirely nonlinguistic form (as in infants and animals). However, there are a number of reasons why these claims appear to me to be spurious and indeed, somewhat curious. Leaving aside for one moment the question of whether FP can be plausibly construed as a theory, it becomes very difficult, once these alleged shortcomings on the part of FP are scanned, to see precisely what theory Churchland takes FP to be. It appears as a very suspicious fact at this point, that Churchland has at no point set forth rigorously a supposed theory called "FP". Until we know what theory FP is, what claims are (supposedly) made by FP, we cannot say whether the areas of ignorance documented by Churchland are actually shortcomings of FP. As should already by clear, Churchland has certainly spoken at points as though what is essential to FP is the conceptual holism of the intentional categories, with its intrinsic connections with action.

But, should Churchland actually be correct (despite Searle and myself) that there is something which can be called "folk psychology", and that it is a theory, then if this supposed theory is merely this intentional holism, then it is difficult to see how the areas of ignorance which he has cited are actually shortcomings

of the theory in question. Doubtless there are a lot of things about and related to the mental about which very little is known; there is indeed ignorance, largely, in the areas mentioned by Churchland. However, again assuming (I stress) that the intentional categories and their essential connections with action are only a theory, then it is not at all clear to me how such a theory is put in jeopardy by the lack of insight in the areas documented by Churchland. For instance, should it ultimately only be some sort of theory (though I think it isn't) that the explanation of a man's climbing a ladder is that he believed that by doing so, he could attain his objective of retrieving his hat from the roof, it is not clear to me how such a theory would be threatened by, for instance, the fact that we don't know a great deal about mental illness. Why need a "theory" of intentional categories explain everything within its purview in order to be taken seriously as an explanation of anything? We can surely see this by comparing the comparable, since it seems to me to be the case in a great many areas that, whilst there may be unsolved problems at the extremities of our concerns, this does not entail that these aren't more mainstream cases within the domain about which we can feel sure: for instance, physicists may still be unsure, ultimately, about whether light is a wave or a particle, but this problem need not throw everything that they hitherto thought they know about light into jeopardy. Similarly, the fact that a cure for a cancer still seems some way off does not mean that everything which physiologists thought they had established must go by the board, nor even more specific claims such as (say) that smoking increases the risk of cancer. In ethics, whilst debates may rage on issues such as nuclear weapons, capital punishment and abortion, this does not mean that we need be thrust into scepticism about whether (say) kicking an innocent old person at a bus stop out of frustration that one has missed the bus, is morally defensible. And so on. The point is, I hope, clear by now. Undoubted cases of problems within the domain of the psychological don't entail the entire collapse of everything we thought we knew about the psychological. And, as I've already said, one thing I think that we do know here is that there are intentional states, some of which (e.g. purposes) are essentially linked to

action, and unlike Dennett and Churchland, I do not see this as in any sense a theory, as some sort of "stance" or conceptual holism which may turn out to have some explanatory efficacy.

Indeed, the later remarks of Churchland in attempted disparagement of the "theory" of FP turn out, on inspection, to be most revealing of all, I think. This is the supposed "stagnation" of FP, and its alleged failure to integrate successfully with theories in the physical sciences. On the "stagnation" theme:

> ...both the content and the success of FP have not advanced sensibly in two or three thousand years. The FP of the Greeks is essentially the FP we use today...one must query the integrity of its basic categories.

And, on the "poor integration" theme:

> ...the greatest theoretical synthesis in the history of the human race is currently on our hands, and parts of it already provide searching descriptions and explanations of human sensory input, neural activity, and motor control. But FP is no part of this growing synthesis. Its intentional categories stand magnificently alone, without visible prospect of reduction to that larger corpus.

It should not by now be difficult to imagine why I think the correct replies to Churchland here support the conclusions he doesn't want to entertain. Churchland doesn't really seem to wonder why this supposed theory called "folk psychology" appears to have been "stagnant" for so long, or why it appears to be so poorly integrated with other established (physical) theories. The answers, I would say, are that there is no <u>theory</u> called "folk psychology", that intentional categories describe (when applied to humans) unambiguously real phenomena in the world which in some cases have essential links with action, and that the intentional categories <u>must</u> stand alone, since they <u>could</u> not, by their nature, be integrated with physical categories. This is, in fact, the correct

account, I believe, of what Churchland views as shortcomings, and damning shortcomings, of the "theory", "folk psychology". So these alleged shortcomings are actually no accident: they are a consequence of, and evidence for, what I have been claiming all along. Intentional states, for instance, have not altered for thousands of years, in that they have been both genuine phenomena and constituted essentially by their directedness, their representing something other than themselves, this feature being precisely the reason why intentional categories could not be a part of a theoretical synthesis of physical categories. Not only are intentional states perfectly real phenomena, but they are essentially constituted by this representative content, and it is this essential feature which no level of sophistication of explanation within the neurosciences could possibly get a hold of. That is, no amount and complexity of explanation involving neuron firings, nerve cells, or whatever, would mention content of my intentional state, my belief (say) "that it is raining outside". These are the reasons why the intentional categories stand alone, it is no shortcoming on the part of the intentional, but an essential feature of real phenomena. (And I'm not sure why Churchland makes such an issue out of the existence of nonlinguistic intentional states - I think it indisputable that they exist, and exactly the same kind of remarks would apply: no degree of physical explanation could grasp the essential nature of nonlinguistic intentional states either). It can be no surprise altogether, therefore, that in the section "Beyond Folk Psychology", Churchland gives no hint as to how we could make this fundamental step from neural dynamics to the intentional, i.e. that outside of the brain which is represented.

Indeed, it ought to be remembered that eliminativism is itself a theory, and yet it seems to eliminate just those intentional states which are necessary to the formulation of any theory. We must pause at this point and ask what we know about the whole notion of a theory. A theory is surely something which it is possible to believe, accept, doubt, understand, etc., believing, accepting, doubting, understanding and the like being things eliminativism is ultimately committed to denying. How, for instance, could anyone be

coherently said to accept eliminativism if it were actually true? A theory is surely about something other than itself, it has meaning, but this also would seem to look very tricky if eliminativism were true. And what is the process of actually arriving at any theory in the first place? The things which Churchland would seem to be doing in "Eliminative Materialism and the Propositional Attitudes", and which, it seems, he would have to have done before he could have written it, are surely things he would be forced to deny if eliminativism were true e.g. thinking, researching, gathering evidence, reassuring, doubting, inferring, believing, etc. And isn't it self-defeating to say that one believes that there are no beliefs?

It seems very difficult to see therefore, how a determinism such as Ginet's H could be made to look more plausible by attempted downgrading of the entire intentional realm. I hope I have successfully argued that such a project is doomed to fail. As Nagel says, physicalism is a kind of idealism, an idealism of restricted objectivity, i.e. turning on the belief that the real features of the world must be accessible to the physical sciences. But, from what I have said, it would seem that the reality which is intentional states, with their essential feature of directedness, cannot be captured by the physical sciences. As Nietzsche says, an essentially mechanical world would be an essentially meaningless world - we simply could not get to the likes of beliefs, desires, intentions, purposes, projects and the likes with only physical categories.

What I have said does not falsify determinism; despite the full-blooded reality and irreducibility of the intentional, I have said nothing which precludes it from being (say) the function of a mere basic neurophysiological, the kind of which has already been talked of. Indeed, I think it worth pausing to re-trace some steps. I have been concerned in this part specifically with eliminative materialism, but I am only concerned with eliminative materialism at all because of the consequences it could have for a discussion of free will. Eliminative materialism, remember, if

true, would make it at least seem much more plausible that determinism is true, since that which seems prima facie, to make determinism most difficult to accept, i.e. the intentional categories, would entirely disappear. The world would be described exhaustively in the physical categories. Eliminative materialism, I concluded, is untenable. However, it must be appreciated that the falsity of eliminative materialism hardly entails either compatibilism or free will. I will very soon be elucidating Van Inwagen's (M) hypothesis and, going back to what I earlier suggested as useful rules of thumb, the falsity of eliminative materialism is perfectly compatible with the truth of (M), a hypothesis which, if true, would kill off any notion of free will; nor does it seem to do anything to invalidate Van Inwagen's First Formal Argument. It is quite consistent to believe eliminativism false, (M) true, Van Inwagen's First Formal Argument sound, and no free will. The significant point of impact here is that the falsity of eliminativism means that one of the conditions of any belief that we sometimes act freely, i.e. the intentional, cannot be thought of as some bogus or inauthentic alleged feature of the world, the intentional is a straightforward real feature of the world. However, I repeat that the reality of the intentional is only a condition (a necessary one) of the existence of free will; it is not sufficient. The falsity of eliminativism only means that an attempt at disparaging the authenticity of the intentional fails, and consequently an attempt at denying a necessary condition of the existence of free will fails.

Before I go on to consider whether it is reasonable to suppose that remaining conditions of free will are satisfied, I want to deal with something else. This is what can be called 'empirical' compatibilism. I have already looked at compatibilist theories which allow that something like Ginet's H may be true, but hold that we can rediscover our freedom in the sacred realm of intentional. I concluded that it can only be by varyingly transparent cheap tricks that such claims can seem plausible. I have also just looked (in Eliminiative Materialism) at a theory which, if true, would help facilitate the truth of physical determinism by downgrading the

impact of the 'given' intentional. I concluded that this could not be sustained either, that the intentional is here to stay. Empirical compatibilism, like 'language strata' compatibilism, allows that something like H may be true, but holds that worries about freedom can nevertheless be shown to be ungrounded simply by becoming aware of conditions of power-ascription. I will conclude that this is ultimately just another compatibilist cheap trick, but it does possess an intuitive appeal and has to be taken seriously.

Notes

1. In *Brainstorms*

6. EMPIRICAL COMPATIBILISM: MORE CHEAP TRICKS

Away at the beginning, I mentioned Keith Lehrer, M.R. Ayers, and J.L. Austin as examples of contributors who, by some means or other, wished to uphold human freedom in the face of the apparent threat of determinism but who (I wanted to claim) can all be seen to be operating with an ultimately futile notion of capacity, which they don't appear to recognise as a problem. A bit later, I also mentioned Harry G. Frankfurt as an example of someone who has tried to purge causality of (at any rate, what I see as) its full-blooded centrality and significance within the problem of free will. In this part, I wish to expand on this.

A cursory reading of the offerings of the people concerned could leave one with the impression that there are significant differences between their respective positions on free will e.g. Lehrer's position in "An Empirical Disproof of Determinism?" is explicitly compatibilist, Frankfurt's position in "Freedom of the Will and the Concept of a Person", whilst less explicit on the matter, is also compatibilist, whilst the mere title of M.R. Ayers' book, The Refutation of Determinism, would seem to suggest that he is, first and foremost libertarian and therefore needn't really concern himself with the question of compatibilism and incompatibilism. However, I would wish to regard those contributors as tied together with regard to what is significant; what is common to them, is, I think, of more consequence than what separates them. What seems to me to link them essentially, is the inclination to ultimately relegate or downgrade causality (with all its alleged importance) with regard to problems of human freedom, on the grounds that the real, substantive issues here can be settled by focusing on much more ordinary and mundane features e.g. the criteria we use to ascribe a power of capacity to someone or (more in Frankfurt's case) the psychological constitution of the kind of person who can be regarded as free. I will now devote some attention to these kind of efforts and, whilst I will recognise that they contain some genuine, valuable insight, and that they help facilitate the clearing away of a lot of dust within the area of free will, I will be arguing that

they are ultimately futile as regards allaying the worries about human freedom which are generated by the prospect of the truth of determinism such as that floated by Ginet or Honderich. (In a previous section, I spoke of Dennett as someone who attempted to divest causality of its ultimate importance, but I do not throw him in with those mentioned here, since his compatibilism is of a significantly different kind. Dennett's position in the face of the determinist threat is, as I've already mentioned, in alignment with that famously propounded by Strawson in "Freedom and Resentment", a position which I will specifically criticise later. The 'empirical' compatibilists, on the other hand, hold that last-gasp saving measures such as those attempted by Strawson are really unnecessary, that we can take the determinist threat at face value and quell it fairly painlessly).

Let me turn now, specially, to Keith Lehrer. I have already noted that Lehrer is explicitly compatibilist. And, in fact, this is really the whole point of his framing the title of his essay in the form of a question, i.e. he claims to have demonstrated, on the one hand, that we can act in a way other than how we do, in fact, act but he wants to ask, on the other hand, whether the success alone of this particular demonstration would falsify determinism ("An Empirical Disposal of Determinism?" - emphasis mine). And indeed, he goes on to argue that it doesn't, i.e. that we can know that someone could have done otherwise, and all our actions may yet be causally determined, indeed, ancestorially determined (Lehrer's expression), this being that the conditions which determine behaviour are themselves determined, etc. The significant similarities that the thesis of universal causal and ancestorial determination has with Ginet's 'H' hypothesis are I expect, plain to see. However, whilst it seemed, after all was said and done, to be an outcome of H that no one could ever act in any way other than how they did, in fact, act, Lehrer (as I've said) is wanting to claim that even the truth of the thesis of universal causal and ancestorial determination does not prevent us from knowing that we could have done other than what we did, in fact, do. This demands examination.

According to Lehrer, we can know empirically that a person could have done otherwise. It is not, he says, merely a matter of seeing him do something at some other time which would justify our claim to know that he can do it at the time when we do not see him do it, but of seeing him do it when certain other epistemic conditions are satisfied:

(a) Temporal propinquity: e.g. if I saw a man perform forty push-ups twenty years ago and have not seen him do it since, that would hardly justify my claim to know now that he could do it. On the other hand, if I saw him do it yesterday, my claim would have much greater merit. The factor of temporal propinquity would, it seems, function in varying ways depending on the particular case, but I think it is a familiar and uncomplicated enough factor for me not to have to labour the point here.

(b) Circumstantial variety: the greater the variety of circumstances under which we have seen the person perform an action, the more justified we are in claiming to know that he can perform it. There would, I suppose be cases for qualification here and there, but again, I don't feel this need be laboured.

(c) Agent similarity: e.g. if we see a man lift a two-hundred pound weight, and he subsequently breaks his arm, our having seen him lift the weight is surely not good evidence that he can do it, now that his arm is broken. Thus, the greater the similarity of the condition of the agent, at the time when we see him perform the action, to the condition of the agent at the time at which we claim that he can perform it, the greater the justification of our claim. Again however, to some extent this condition could be formulated as a condition of variety rather than one of similarity, i.e. if we have seen the agent perform an action at times when his condition has varied greatly, then even though the condition of the agent at the time at which it is claimed that he can perform the action is

quite different from what it was when we saw him perform it, the claim might, nevertheless, be fairly well justified.

(d) Simple frequency: Other conditions aside, the more frequently we have seen a person perform an action, the more justified we are in claiming to know that he can perform the action when we do not see him perform it.

Lehrer goes on to say that if all of these conditions are very well satisfied with respect to any action, we possess sufficient empirical evidence to support the hypothesis that a person can perform it, and in the absence of any evidence to the contrary, we are certainly justified in claiming to know that the hypothesis is true. And indeed (Lehrer concludes), by the usual canons of inductive evidence, our evidence is excellent. He goes on and asks us to imagine an expirement which would enable us to obtain such evidence. We are to imagine a subject who is normal in every way, and fabricate an expirement to investigate when he can, and when he cannot, lift an arm. We might first instruct him to lift his arm whenever we tell him to, and see that he does this. We might then (the experiment goes on) tell him to lift his arm whenever we tell him not to and see that he does this. We might then tell him to heed or not to heed our instructions, and see that he sometimes lifts his arm when we tell him to, and sometimes does not, and that he sometimes lifts his arm when we tell him not to, and that he sometimes does not. We could (Lehrer continues) then run the experiment under a variety of circumstances, indoors and outdoors, under stress and under relaxed conditions, with a weight attached to his foot and without impediments, etc. We might also keep careful records of the subject throughout, and finally, vary his condition by drugs, hypnotism, etc. And suppose that we then instruct the subject to heed or not heed our instructions, and ensure that both he and the conditions are those we have found most propitious for arm-lifting. Moreover, suppose that we watch him lift his arm, then avert our eyes for a moment, and subsequently see him lift his arm again. In this case, notes Lehrer, the conditions of temporal propinquity, circumstantial variety, agent similarity, and simple

frequency would be satisfied. According to the thesis we have already seen Lehrer put forward, we would consequently have sufficient empirical evidence to support the hypothesis that the agent could have lifted his arm during that brief period when we did not see him lift his arm, and consequently we would be justified in claiming to know that the hypothesis is true. Indeed, this claim would be justified whether or not the agent lifted his arm at the time in question and indeed, would be justified even if we knew that he did not lift it. If (Lehrer claims) we are able to rule out the hypothesis that the agent tried and failed, and if the condition of the agent as well as the circumstances in which he is placed are those we have found to be most favourable for arm-lifting, then the mere fact that he does not lift his arm would not support the hypothesis that he cannot lift it.

So Lehrer has, it seems, demonstrated that, regardless of the truth of determinism, we can know that a person can do otherwise. But I have already agreed with Ginet and Van Inwagen that, if H is true, then a person cannot behave in any way other than he, in fact, does. Are those two hypotheses irrevocably at odds, and if so, which one is correct? If H is true, can we, or can't we, do other than what we do? It has to be said right away that the Lehrer enterprise that I have just presented does seem to go through so smoothly that it is difficult not to find it appealing. However, after exploring ways in which it may be helped to look convincing i.e. moves which may help to make it look reasonable to say that we can do otherwise in the face of the truth of H, I will be forced to conclude that this appeal is ultimately illusory.

We can be pointed towards the source of this illusion I feel, by considering the different usages of the word 'can'. This is because I think that one could only be deceived by Lehrer's compatibilist manoeuvrings if one invests too much in the undoubted fact that there are significantly different uses of 'can' (or 'could have done'). I will clarify, at some length. There is the usage of 'can' which has one say that what is the outcome of a set of sufficient conditions cannot fail to happen, a usage which, were

Ginet's 'H' hypothesis or Honderich's ('One') determinism true, would always render false the claim that I can do something other than what I, in fact, do. However, Honderich himself, in 'One Determinism', recognises another usage of 'can', according to which what does not happen can happen, even if determinism is true:

> We sometimes say that B can happen in a certain situation and mean, essentially, that B will happen if some further condition is satisfied. This is not, as in the first usage, a denial that not - B will be caused to happen. Indeed, something that cannot happen, in the first usage, will be something that can happen in the second usage.

It should not be difficult to see what Honderich is getting at here. In what Honderich cites as the second usage of 'can', 'can' is logically linked to some set of conditions. 'Can' reduces here to 'will if...'. This is, I think, uncontroversial enough and there can be little doubt that there are both these usages of 'can', as documented by Honderich. This can be reinforced even by fairly banal examples. The heavy rain may have been so unpleasant that I coughed up the money for a bus ride and, whilst sitting on the bus, I lament that I could have avoided this, i.e. if I had remembered my umbrella. The rain caused me (with the proviso of various background conditions perhaps) to do one thing (and which in the first sense of 'can', I could not have avoided doing), but I would have done a different thing had a particular condition been satisfied (this providing the second sense of 'can', according to which I could have walked, instead of taking the bus). So let us, for the moment, postulate that the sense of 'can' or 'could have', which Lehrer is employing, according to which we can act otherwise even if determinism is true, is some form of the second sense. It clearly cannot be the first sense (since the mere truth of determinism, which Lehrer does not disclaim, would preclude this), but perhaps there could be a variant of the second sense which Lehrer could be putting forward, which could give plausible content to the notion of genuinely open possibilities of action, in the face of determinism. Let us see if it is possible to effect this sort of

marriage between Ginet's 'H' hypothesis and Lehrer's claim that we can know empirically that we can act other than how we do act. Lehrer, remember believes that something like H could well be true, so what we must do here is examine what content Lehrer's specific thesis regarding our doing otherwise could actually have. The attempt I will make at displaying compatibility between Ginet's H hypothesis and Lehrer's claim that we can know empirically that we can do something at a time when we are not doing it will be, I think, ultimately revealing regarding Lehrer's compatibilism, and the specific sense of 'can' or 'could have' which Lehrer is employing.

It should be clear by now, having noted what is Honderich's second sense of 'can' that in the face of the truth of H, it is yet possible to attach some sense to the notion of a person's being able to act in a way other than how they do act. However, we must be very clear as to precisely what this sense is. That is, if H is true, we could say (sticking to Ginet's original variables) that when a person did B and could have done otherwise, we mean that he would have done otherwise if A_1 had not occured, and some other circumstance (say, D_1) had been the first of the antecedent circumstances. If that had happened (we could clarify), instead of A_2 being contingently necessitated, D_2 would have been contingently necessitated, which in turn, would have contingently necessitated D_3 (instead of the A_3 which actually happened)... and D_{n-1} would have contingently necessitated D_n, which would have contingently necessitated (say) C (instead of the B which actually occured). The fact of the matter is (we could finish) it was A_1, and not D_1 that occured, and that was why he did B and not C. That is:

(1) A_1 contingently necessitates A_2, A_2 contingently necessitates A_3,..., A_{n-1} contingently necessitates A_n, A_n contingently necessitates B
(and of course, the man had no choice as to whether A_1 would occur), and

(2) D_1 contingently necessitates D_2, D_2 contingently necessitates

D_3, ..., D_{n-1} contingently necessitates D_n, D_n contingently necessitates C
(the man had no choice as to whether D_1 would occur).

So perhaps we may be able to say that, whilst we will use Lehrer's four conditions by way of establishing whether a person can do something that he is not doing, whether or not he actually does it is itself ancestorially determined by the kind of factors H includes. Again, we could employ Lehrer's four criteria and say, at a time when a person is not doing L: "He can do L", and then refer to H to discover that he would be doing L just now if X_1, which contingently necessitates it, had occcured, and that the reason he is doing M just now and not doing L is that Y_1, which contingently necessitates M, in fact occured, and X_1 did not occur.

However a shrewd reading of what has just gone should move one towards what I think is really the punch-line. And that is that, whilst I have distinguished, a la Honderich, between two different usages of 'can', it is now clear that, when Lehrer claims that we can know empirically that someone can do otherwise, he is not using 'can' in either of those two senses. Since Lehrer is not denying determinism, he is clearly not using 'can' in the first sense specified. But what of the second sense? With this, to say that a person can do otherwise, is to say that under a certain set of conditions different to those which held, he <u>would</u> have done otherwise. Before going on however, it may well be worth pausing and asking what import this itself has. That is, if Ginet's H hypothesis is true, then no one ever has any part anyway in determining which set of conditions (e.g. 'A_1' or 'D_1') happen to be true of him at any given time, and consequently no part in deciding how he ultimately behaves. So, if 'H' is true, it is not at all clear how this second usage of 'can' does anything at all for the cause of freedom and indeed this feature, as well as the ultimate fate of Lehrer's efforts, should be especially clear soon, when I document Van Inwagen's Second Formal Argument. That is specifically why I spoke earlier of <u>variants</u> of this second usage of 'can', since its mere existence clearly secures nothing for freedom; the question for freedom seems to turn on what the conditions are which

fill the gap in 'will if...'. It cannot be just anything, it has to be something which facilitates my getting into my action somehow (and, of course, the position of the conditional analysis of power within the history of free will/determinism, e.g.'I can' means 'I will, if I choose', is itself well-known, conditionalism being something I will properly consider later). However, as I've already suggested, this doesn't really seem itself to be a problem for Lehrer and, as I've also hinted, Lehrer's problem seems to me to be much worse than this. On Lehrer's analysis, to say that someone could have done otherwise is not to say that he would have done otherwise under a certain specified set of antecedent conditions which didn't in fact hold (any set). It seems, on the contrary, that all one would be committing one's self to is the satisfaction of Lehrer's four conditions of power-ascription i.e. absolutely not to the claim that there is a certain set of conditions under which someone would have done otherwise. Those conditions could go on being satisfied forever, with the 'other' action never occuring. The man (in Lehrer's example) may never lift his arm, despite the persistent satisfaction of the conditions which, according to Lehrer, allows us to say that he can lift it. That is why (the other problem I have posed about this being waived for the moment), (1) and (2), or anything significantly like them are, by their nature, crucially incorrect as an account of Lehrer's claims for his example, as well as his more general claims about powers. (1) and (2) display our second usage of 'can' but as I've said, Lehrer's usage of 'can' is clearly not this one. It is therefore appropriate to ask what Lehrer is up to. What significance do his efforts have for free will/determinism? Given that Lehrer accepts that something like Ginet's 'H' hypothesis may be true (and that his use of 'can' is not our second usage), what can his claim, that we can do other than what we do, actually amount to?. Remember that Lehrer's defence of our ability to do otherwise consists, basically, of a recounting of our fairly ordinary power-ascription procedures. According to Lehrer, by employing these banal procedures in the given situation, we will come to know fairly unproblematically, whether we are entitled to say that someone can or cannot do a particular thing which he is not now doing. If one views the matter as Lehrer seems

to do, i.e. if one believes that questions of powers, or at least, the only questions regarding powers that we need be interested in, are entirely settled in this simple, empirical way, then free will/determinism seems a sadly vacuous issue, on which a lot of pointless energy would seem to have been wasted. Quite regardless of whether the suggestions of the likes of Ginet and Honderich are correct, we know that we could go on indefinitely, employing those simple, empirical power-criteria, criteria which (Lehrer thinks) yield the conclusion that we can often do something at a time when we are not doing it. According to this, even if determinism is true, the attendant worry regarding our freedom is bogus, and a simple reminder to ourselves of our ordinary procedures will allay the worry. However, such claims seem to me far from being implausible, to be actually a side-stepping, an evasion of the entire issue, the kind of manoeuvre which Nagel calls in his introduction to The View from Nowhere, a "cognitive wish-fulfilment". As Nagel himself recognises, much recent philosophy has been sadly infected with the tendency to construe the traditional problems of the subject as being much less genuine or difficult than they are, and Lehrer's enterprise here seems to me to be a case in point. If all worries regarding determinism were as groundless as the empirical analysis which Lehrer (and some others) suggest, then it would be very difficult to see why the question has occupied the minds of so many people over such a long period. It is surely worth saying that the likes of Ginet and Honderich (and Van Inwagen, who will be especially noted soon) would have little problem recognising that we do, in fact, employ criteria of power-ascription in a fashion something like that documented by Lehrer e.g. we do tend to say the likes of 'I can do twenty press-ups' on the grounds that I did it yesterday, there having been no deterioration in my fitness since then, etc. I don't think, however, that this fact is any better than banal, as regards contributing to the free will/determinism issue. If Lehrer is really to make any serious contribution to free will/determinism, it seems to me incumbent upon him to move beyond what is really only a descriptive exercise, a piece of linguistic commentary almost, and ask another set of questions e.g. what do I mean exactly when I say

that I can do twenty press-ups? What is the connection between this claim and the power-ascription criteria I use? How could it square with the truth of determinism? It is on these questions that I find comments made by Van Inwagen in An Essay on Free Will especially illuminating. In fact, I think that he is more than helpful on this issue at several junctures in the book, and in somewhat differing contexts. Because of the deceptively appealing look of Lehrer's compatibilism, and the consequent need to reinforce the illusory grounding of this appeal as strongly and precisely as possible, I will cite several strands within Van Inwagen's approach which I consider pertinent. Early on in the book (1.4) Van Inwagen distinguishes between a skill, accomplishment, or general ability, on the one hand and, on the other, the power to exercise it on a given occasion. It seems to me that Lehrer has, all too covertly and with calamitous consequences for the discussion of free will/determinism, managed to conflate what it is essential to distinguish. I have already talked of different usages of 'can' and in fairness to Lehrer, his conflation has been facilitated somewhat by something Van Inwagen recognises, namely that 'can' is clearly used in both kinds of situation. However, as Van Inwagen (crucially) recognises:

> It is plain that the 'can' that figures in discussion of free will and determinism is not the 'can' of skill: the thesis of determinism may or may not be relevant to the question whether someone on a particular occasion can or cannot speak French; it is certainly irrelevant whether that person is a French-speaker.

I hope that my major point is becoming fairly clear by now, and is looking defensible. Despite the warning we have just seen Van Inwagen give, it looks very much as though Lehrer has allowed the 'can' of skill to slip into his discussion of free will and determinism. As with Van Inwagen's example of the French-speaker (or otherwise), the thesis of determinism is indeed irrelevant to the question of whether I satisfy various kinds of conditions (those loosely documented by Lehrer) associated with the performing of

twenty press-ups; Lehrer is correct about this. However, the thesis of determinism may well be relevant as to whether I can, on a particular occasion, perform twenty press-ups. As I've already indicated by my general disparaging of Lehrer's contribution, I am in agreement with Van Inwagen that it is this point which is important with regard to free will and determinism. Lehrer seems unable to recognise this. In fact, I believe that even more penetrating comments on this are made by Van Inwagen later in his book, in Chapter IV, "Three Arguments for Compatibilism", where he specifically discusses the Paradigm Case argument. I think it reasonable to regard Lehrer's position as a variant of this argument. The classical statement of it, of course, is due to Antony Flew, and it is useful, I feel, to recount a part of it at this point.

There are various words and phrases we use in ascribing free activity to people besides the obvious 'acted freely' and 'did it of his own free will', there are such phrases as 'could have done otherwise', 'had a choice about what she did', 'had alternatives', and 'could have helped doing what he did'. We learn these phrases by watching people apply them in concrete situations in everyday life, just as we learn, for example, colour words. These concrete situations serve as paradigms for the application of these words: the words mean things of that sort. Therefore they must apply to something; they must apply at least to the paradigmatic objects or situations. As Van Inwagen recognises this is not, of itself, an argument for the compatibility of free will and determinism; rather it is an argument for the existence of free will. However, it is when we note one of the arguments put forward by Flew to supplement this one, to show that free will (as well as existing) is compatible with determinism, that we ought to be clearly reminded of Lehrer's enterprise. That is, when we carefully investigate the paradigm cases of free action, we find that their common feature is just this: we apply the word 'free' to a person's act just in the case that "if he had chosen to do otherwise he would have been able to do so; that there were alternatives, within the capacity of one of his physical strength, of his IQ, with his knowledge, and open to a

person in his situation". And, of course, a person's act may have this feature whether or not determinism is true. Indeed, later, Van Inwagen puts this general compatibilist argument yet another way:

> Our everyday ascriptions of the ability to act otherwise make no reference to determinism. That is, we do not find out whether an agent's act was undetermined by past events in order to find out whether he could have acted otherwise. Therefore, the thesis normally expressed by 'he could have acted otherwise' does not entail the falsity of determinism.

Although it is not really of great importance for my purposes, it strikes me as slightly odd that, despite Van Inwagen's displaying acquaintance with Lehrer's article a moment later, he reports that he is unable to find a clear example of this (above) argument in print. Lehrer's argument strikes me as a very clear example of it. However, the main thrust remains unaffected. As Mackie instructs in his essay "Responsibility and Language",[1] it is a mistake to start with the assumption that spheres of thought (and practice, I would add) are separable and do not really conflict, this being a mistake into which a linguistic philosopher is particularly liable to fall (though, as I said earlier, irony of ironies, Mackie seems to make precisely this mistake vis à vis free will/determinism and moral responsibility). Indeed part of the objective of Van Inwagen's Second Formal Argument (which I will document soon), is to ensure against such an illusion by talking (unlike Lehrer and Flew) of free will and determinism in the self-same vocabulary (of "possible worlds"). I am myself wholly in agreement with Van Inwagen (and in disagreement with Lehrer and Flew) that it is false that the thesis normally expressed by 'he could have acted otherwise' does not entail the falsity of any general thesis to which our everyday ascriptions of the ability to act otherwise make no reference. This point is brought out most shrewdly by Van Inwagen when he postulates an imaginary state of affairs which I, in fact, believe that <u>all</u> empirical (or "over-empirical") contributors to the free will/determinism issue should note. I have pre-empted this, specifically in my discussion of the approach of Davidson, Dennett

and Mackie to issues in free will/determinism (and my own remarks must have the same consequences for Strawson, and, as I note in the next section, M.R. Ayers). Whilst it is, as Van Inwagen concedes, fanciful, it is also as he says, logically adequate, and I believe that it is of the greatest importance with regard to the specific point presently being discussed. Van Inwagen calls this proposition '(M)':

> (M) When any human being is born, the Martians implant in his brain a tiny device - one that is undetectable by any observational technique we have at our disposal, though it is not <u>in principle</u> undetectable - which contains a "program" for that person's entire life: whenever that person must make a decision, the device <u>causes</u> him to decide one way or the other according to the requirements of a table of instructions that were incorporated into the structure of the device before that person was conceived.

I hope that it is not, by now difficult to see that the crucial point about (M) is that it is consistent with all our observations (and practices), but has the consequence (despite Lehrer and Flew) that no one can do otherwise than he does. As Van Inwagen is himself aware, someone might object that (M) is not in fact consistent with our observations, since we can normally "feel" our decisions "flowing" naturally from our desires and our beliefs; but if (M) were true (the objection runs), we should "feel" ourselves being interfered with. But of course, to meet this objection we need only suppose that the Martian device causes us to have desires and beliefs appropriate to the decisions it will cause us to make. If the Paradigm Case argument were valid, then it would follow that (M) was compatible with free will. If (M) were true, the world would look the way it in fact looks, and our linguistic practices would be the same - or at least (as Van Inwagen points out), we should emit the same sounds in the same situations, and our production of these sounds and their reception by our audience would be accompanied by the same internal sensations. If the Paradigm Case argument were correct, the extension of the term 'free' could

be just what it in fact is even if (M) were true, since (M) does not require any observable features of the world to be different from what they are in actuality. But (M) obviously does entail that no one can act otherwise than he does. As Van Inwagen says, if we should discover that some particular person acted as he did because a Martian device, implanted in his brain at the moment of his birth, had caused all his decisions, then we should hardly want to say that he had free will, that he could have helped what he did, that he had any choice about the way he acted, or that he could ever have done otherwise. And if we discovered that everyone was directed by a Martian device, then we should have to make these judgements about everyone. As I suggested earlier, when I spoke specifically about them, this argument can be issued to good effect against 'language strata' theorists such as Davidson. Van Inwagen is perfectly well aware that there are all sorts of important differences between determinism and (M), which we can quite easily appreciate, I think, by setting (M) alongside the formulations of determinism attempted by not only Van Inwagen, but also Ginet and Honderich. For instance, (M) entails that each human being's dispositions to act are chosen for him by a non-human intelligence, and determinism does not. However, this is not really what is important: what is important is that the fact that we can construct propositions such as (M) shows that it is at least possible that certain propositions that we know of but have not ourselves constructed are also consistent with our observations but inconsistent with the free will thesis. And, of course, for incompatibilists such as Ginet, Honderich, Van Inwagen and myself, determinism is such a proposition. As Van Inwagen suggests, this conclusion can hardly be shocking, since determinism is a very general thesis, essentially involving concepts (like 'law of nature') that are intimately connected with the concept of ability (since no one is able to change the laws of nature). Indeed, I want to reinforce this connection (and then incompatibilist conclusion), by considering Van Inwagen's Second Formal Argument in favour of incompatibilism. Before doing even this however, let us reflect on Lehrer's efforts vis a vis what I have called the two useful "rules of thumb". Does anything said by Lehrer discredit Van Inwagen's First Formal

Argument? I don't see that this is the case. And, having now seen (M), we can ask whether anything said by Lehrer would make it possible to save free will, if (M)) were true? Again, I don't see that Lehrer has done this.

It is, I think, especially salutary to introduce Van Inwagen's Second Formal Argument at this point. As I've already hinted, this is because so much confusion (and, alas, compatibilism) has been encouraged by differing vocabularies within the debate. We have already seen the kind of vocabulary which the H hypothesis or a law of nature is going to be expressed in and this, of course, seems of an altogether different type from that which will be used, for instance, in attributing skills to persons. We can recall Mackie's intellectually invaluable point again, and ensure that free will and determinism occupy the same territory. Indeed, Van Inwagen anticipates the charge of artificiality against the language of his Second Formal Argument, with the rebuttal that the "artificiality" has a good excuse: if we are to investigate the conceptual relations between free will and determinism, it is hardly to be supposed that we shall succeed if the vocabulary we use to state the thesis of determinism and the vocabulary we use to state the thesis of free will have no elements in common. Therefore, if determinism is formulated as a thesis about possible worlds (or about porpositions), the best plan would seem to be to try to formulate the free will thesis as a thesis about possible worlds (or propositions). I agree with Van Inwagen that the concept of a possible world is an extremely useful one and, like him, find it difficult to understand the number of otherwise intellectually responsible people who have amused themselves by sneering at it. It is worthwhile therefore, to cite some of the remarks made by Van Inwagen in defence of this notion, a defence which Van Inwagen admits is largely cribbed from Alvin Plantinga's <u>The Nature of Necessity</u>.

Possible worlds are members of the class of <u>ways things might be</u> or <u>possible ways things might be arranged</u> or, simply, possibilities. We quantify over such objects when we say (for

instance) "There are three ways in which the Earth could be destroyed", or "There are still a few possibilities that we haven't investigated" (Van Inwagen's examples). A possibility includes a second possibility if it is impossible, for the former to be realised and the latter unrealised (and every possibility includes itself). A possibility precludes a second possibility if it is impossible for them both to be realised (no possibility precludes itself). Thus, the possibility that Socrates teach Plato includes the possibility that 2+2=4, the possibility that Socrates exist, the possibility that Plato exist, and the possibility that Socrates teach someone. It precludes the possibility that Socrates fail to exist and the possibility that Socrates teach no one. It neither includes nor precludes the possibility that Socrates teach Aristotle or the possibility that the most famous snub-nosed Greek philosopher teach Plato (courtesy Van Inwagen). There is a good deal more said by Van Inwagen by way of preliminary explication and defence of the vocabulary employed in the Second Formal Argument, but I believe that much of it can be waived without great loss. If only because of the basic limitations of time and space, one must sometimes, in philosophy as in any other discipline, take on board the assumption that a solution is possible to some or other problem, without being sure precisely what the solution is. This hardly means, of course, that one goes around doing it willy-nilly or gratuitously because, beyond a certain domain, it would bring one's entire efforts into disrepute. Where I think it is a defensible strategy is in the case where there already seem to be very good reasons for being sympathetic to some or other general notion, these reasons grounding a confidence that the specific problems generated by the idea to which one does not presently have the solution, do in fact admit of a solution. For instance, I would, in some kinds of context, regard the notion of personal identity over time (continuity of) as such a case. I cannot say with confidence what it is exactly that accounts for what seems to me to be the undoubted fact that I am identical with one Paul Davis who existed ten years ago, and who was then about to enter his fifth session at a secondary school, etc. However, because I think there are compelling reasons for believing in continuity of personal identity over time, I would think it no

great liberty to simply assume that there is a solution to the specific issue of the continuity of my identity over time. Similarly, I think, with the "possible worlds" notion: since (as I've said) I am, like Van Inwagen, generally sympathetic to the belief that the whole idea is useful and defensible, I think that I can be allowed, whilst recognising the existence of a host of highly specific and complex problems generated by the notion, to simply get on, and assume that these problems admit of a solution. (Indeed, frequently of course, the outcome of one's attempted use of the notion further grounds one's confidence that a solution to the outstanding problems can be found.) Van Inwagen himself ends up just saying this kind of thing occasionally. So, before getting down to the actual statement of the Second Formal Argument, I will restrict myself to only two more stage-setting preliminaries, and simply clarify as I go along, should I think it necessary.

A proposition is true at, or in, a given world if that proposition would be true if that world were actual. It is an obvious consequence of this definition that there are objects that exist in more than one world. For instance, there are possible worlds, none of them actual, at which Socrates had a long, straight nose (again let us simply assume a solution to any problem thre may be of "trans-world identity"). One world shares a slice with another if the two worlds are indistinguishable at some instant (I will waive any problem regarding absolute indistinguishability at any instant). The property of being deterministic is informally characterised by Van Inwagen as follows: a world is determinstic if that world itself is the only world that both shares a slice with it and has the same laws of nature it does. With his characteristic shrewdness, Van Inwagen goes on to provide an example which splendidly illuminates this definition. Let W be some possible world that shares with the actual world, A a slice taken at the instant Harold's eye was pierced by a Norman arrow. And let us suppose that in A and W the laws of nature are the same. A and W can plausibly be called 'deterministic' only if they are identical. For suppose they are not identical: let us say that W is one of those worlds in which a thermonuclear war was fought in 1966. If

there is a world that has all the properties we have ascribed to W, it would be odd to say that anything that could reasonably be called 'determinism' is true. The actual world is a world in which a certain situation in 1066 did not precede a thermonuclear war by nine hundred years. But in W, a world having exactly the same laws of nature, precisely the same situation was followed, after nine hundred years, by a thermonuclear war. In other words (Van Inwagen goes on), if our description of W is consistent - that is, if W exists - then, though there was in actuality no thermonuclear war in 1966, such a war was a possibility relative to the laws of nature and the state of the world in 1066. But 'determinism' is the thesis that there are no such alternative possibilities: according to determinism, every world distinct from the actual world either differs from it at every instant, or if it differs from the actual world at only some instants, is governed by a different set of laws of nature. That is, to all intents and purposes, what Van Inwagen, in his Second Formal Argument, has to say about determinism within the vocabulary of possible worlds, and I hope that it is now beginning to look to any sceptics there may have been that there is some substance to the belief that "possible worlds" can do a job. As I've already said, one of the primary objectives of this approach on the part of Van Inwagen is to ensure that we can talk about free will and determinism in the same vocabulary, and I reiterate that I cite it at this juncture precisely that we can be disabused of the compatibilist confusions facilitated by the duplicity of language allowed by the likes of Lehrer and Flew. So it should be clear that we must (and why we must), when we move from determinism to free will, continue to speak in the language of possible worlds. In another locution which more than adequately, I think, captures exactly what is required, Van Inwagen talks of a person's abilities in terms of which worlds he "has access to". Non-actual worlds are unrealised possibilities. Thus "access" talk is a way of organising our talk about unexercised abilities by reference to unrealised possibilities: an unexercised ability is treated to realise some unrealised possibility. For instance (Van Inwagen's example), we translate, 'Napoleon could have defeated Wellington at Waterloo' as 'Napoleon had access to some possible world in which he defeated

Wellington at Waterloo', and we translate 'It is within my power to keep the money I found and I have access to some world in which I return it'. Again, of course, as Van Inwagen is perfectly well aware, there are fringe difficulties, at least, with this idea; there are undoubtedly ways in which it is unrealistic. However, as Van Inwagen points out, there are ways in which the Kinetic Theory of Gases is unrealistic; it ignores certain features of the real world, such as inter-molecular forces. The point of the simplifying assumptions is to lay bare the most important features of an interrelated family of phenomena. Before eventually going on properly with Van Inwagen's Second Formal Argument, I will restrict myself to citing a metaphor he provides for his notions of access to possible worlds, a metaphor which, although laden (as he recognises) with the kind of difficulties I have alluded to, is nevertheless, highly illuminating on the primitive terms on which it is intended to be taken. Consider a man who is walking through an infinite system of branching corridors. He has always been walking and must always keep walking, never stopping and never retracing his steps. He finds that some branches are sealed off by bars and some are not. Frequently he comes to a branching of the corridor from which at least two unbarred branches lead away, and he must make a choice about which to take. We can call a <u>possible world</u> any infinitely long path through the system of corridors that does not cross itself. The <u>actual world</u> is that one path, through the corridors along which the man always has walked, is walking, and always will walk. Those worlds to which the man <u>has access</u> at any given moment are just those infinite paths that do not pass through any barred corridors, and which are continuations of the path-segment along which he has already walked.

Getting back to the free will side of the Second Formal Argument proper, Van Inwagen states the <u>minimal free will thesis</u> (MFT):

> $(\exists x)\ (\exists y)\ (Hxy\ \&\ y \neq A)$, where 'A' is the actual world and 'Hxy' means 'x has access to y'.

MFT tells us only that some person, past or present or future, had, has or will have access to some possible world besides the actual world. MFT is true, for example, if Julius Caesar had access to some world W in which he did not cross the Rubicon, even if no other person, past, present, or future, has access to any world besides 'A', and Caesar himself had access only to W and A (Van Inwagen's example). But if MFT is false, then any more interesting free will thesis is false. And, therefore, if determinism is incompatible with MFT, it is incompatible with any more interesting free will thesis. Van Inwagen is aware that, as it stands, it is not obvious that bare determinism is incompatible with MFT. However, with the augmentation of two seemingly fairly painless assumptions cited by Van Inwagen, it seems that the denial of MFT may be formally deduced from determinism. Van Inwagen calls these specific assumptions "metaphysical assumptions":

MAA $(x)(y)\ (Hxy > SyA)$
MAB $(x)(y)\ (Hxy > NyA)$

where 'SyA' means 'y shares a slice with A (the actual world)', and 'NyA' means 'y is nomologically congruent with A' (or 'y has the same laws of nature as the actual world'). MAA asserts that every world to which any person has access must be indistinguishable from the actual world at some instant. For example, as Van Inwagen puts it, however many possible worlds I have access to, surely they must all be indistinguishable from the actual world at some time in the remote past - say, 10,000 BC, or indeed, any time before I was born. MAA may be regarded as a statement of the familiar principle that no one can change the past. MAB asserts that no person has access to any world in which the laws of nature are different from what they are in the actual world. And this seems undeniable, for no one can render a law of nature false. Laws of nature are not a matter of human choice. And it should hardly require Van Inwagen's finishing formal touches to see that the denial of MFT is derivable from determinism, MAA and MAB. I shall indulge a bit however. If, like Van Inwagen, we formalise 'x is deterministic' (Dx), a la

$Dx = (\exists y)(Sxy) \ \& \ (y)(Syx \ \& \ Nyx > y = x)$, and

represent determinism in the present vocabulary by the formula 'DA', then the argument having 'DA', MAA and MAB as its premises, and the denial of MFT as its conclusion, is the Second Formal Argument. Assume 'Hxy'. Universally instantiate MAA and MAB. Modus Ponens yields 'SyA' and 'NyA'. From these two formulae and the universal instantiation of the second conjunct of the formula that 'DA' abbreviates, we get, by Modus Ponens, 'y = A'. Thus by Conditional Proof and Universal Generalisation, we have '(x)(y)(Hxy > y = A)', which is logically equivalent to the denial of MFT.

The bottom line is this: if determinism is true, then no one has any access to any non-actual world. That is, no one has an ability that may correctly be described as an ability to realise some in fact unrealised possibility. I think that Van Inwagen's Second Argument is not only admirably rigorous (like his First Argument), but brings out into the light exactly what the consequences are, in cash terms, of the truth of determinism, consequences which no amount of huffing and puffing or duplicity on the part of the likes of Lehrer and Flew (or for that matter, Davidson, Dennett, Mackie or Ayers) can alter. Lehrer's empirical criteria (which, as we've seen, are criteria of skill-ascription) are no metaphysical (or moral) knight in shining armour, to save us from what may be the ultimate consequences of our containment in the natural world. As I hope is completely clear by now, what is of importance is that if determinism is true, then there is one world, and one world only (i.e. the actual world) open to us. There seems no way that the maiden can be rescued from the dragon or, at any rate, if there is a way, it is certainly not as Lehrer (or Flew) seem to believe.

If my preceding argument has been sound, then another attempt at downgrading the ultimate importance of determinism vis a vis freedom looks doomed. However, this attempt at marginalising determinism is not entirely exhausted by the specific forms of it which I have so far considered. I have already mentioned Harry G.

Frankfurt in this connection and, whilst I don't intend to elucidate on him until a bit later, I think it worth giving a foretaste of his position. By attempting to exploit a purported distinction between "free will" and "freedom of the will", Frankfurt thinks that he can describe the significant features of someone who "lacks nothing in the way of freedom", whilst remaining entirely neutral with regard to the question of free will and determinism. That is, Frankfurt thinks that determinism could be true, and yet there still be a person who lacks nothing in the way of freedom, since the question of determinism needn't even be relevant. How such a prima facie implausible position can be entertained is something I will document later.

In the meantime, I wish to consider something else I think of relevance to the whole issue of free will/determinism. I have already considered a version of empirical compatibilism, the claim that a fairly straightforward examination of facts about the world and of our ordinary power-ascription procedures reveals that, regardless of the truth of determinism, we do have free will, we can act other than how we do, in fact, act. I argued that this just won't do. However, I now wish to consider something significantly similar, something which could be regarded as the other side of this empirical coin, really. It could itself be called "Empirical Determinism", or more strictly speaking, I suppose "Empirical Determinism & Incompatibilism", since it is the claim that a shrewd analysis of ordinary facts about the world reveals that we cannot, in fact, act in any way other than how we do act. It is worth reiterating at this point that, if Ginet's H hypothesis were true, then the conditions which determine our behaviour (contingently necessitate), regarding which we have no say are what Ginet calls "physiological-cum-environmental". This characterisation may itself seem cumbersome and somewhat curious, but this is not really my concern at the moment. I mention it again because the traditional arguments in favour of empirical determinism/incompatibilism have homed in on physiological and environmental factors as conditions of action, i.e. the suggestion is that a proper appreciation of the importance of physiological and environmental factors, an

appreciation which can be gained without particularly sophisticated or advanced techniques of enquiry, will reveal that we can never in fact, do other than what we do. This really amounts to saying that we can discover that Ginet's H hypothesis is, or something very like it, is true, and that we can discover this with much less investigative agony than Ginet thinks would be necessary to settle things one way or the other. According to this thesis, someone like Ginet really can't see the wood for the trees, since a fairly shrewd insight into fairly ordinary facts evidences something with the same upshot as H, suggests very strongly that physiological and environmental factors leave us unable to act other than how we do, in fact, act. Whilst I agree with the remarks made by Peter Van Inwagen in the preface to his Essay on Free Will, that M.R. Ayers' The Refutation of Determinism, is one of several books supposedly on the topic of determinism which never actually properly addresses the central issues put forward by the likes of Ginet, Honderich and himelf, I think that Ayers does, however, have some sound comments to make on empirical determinism/incompatibilism. Even then, however, Ayers fails to identify and criticise what seem to me to be some very common features of this attack on behalf of determinism/incompatibilism. I don't think it possible to speak with proper penetration about this approach without first talking about what I want to call "explaining away" arguments. I believe that empirical determinism/incompatibilism and "explaining away" arguments are inextricably bound up.

Notes

1. In J.L. Mackie, Persons And Values.

7. "EXPLAINING AWAY" AND EMPIRICAL DETERMINISM/INCOMPATIBILISM

I have already pre-empted this, to some measure, when I spoke of Davidson's Psychological Anomalism, and the status of the psychological with regard to Ginet's H hypothesis. I hope that, by the end of this chapter, the connection between (and coherence of) my remarks at that point, and my remarks now, will be entirely clear.

"Explaining away" arguments proceed, basically, as follows: the thesis (whatever it may be) is put up for grabs and all apparent counter-evidence is explained away in terms of the thesis itself, leaving it very difficult to see what the thesis actually is, making it look as though there can be no evidence against the "thesis". The thesis is eventually revealed, therefore, to be something which is not actually open to falsification at all, but which is, in fact, an a priori assumption which will be used to interpret all data presented. A great many issues which have received serious philosophical treatment have tended to be a prey to this sort of approach, and indeed it has often only been because this approach has permeated the thing that there has appeared to be a serious philosophical problem in the first place. Perhaps the most obvious and celebrated is the question of egoism. This is the claim that human beings always only act out of their own self-interest (and there is, of course, an even stronger claim, entailing the weaker one, that humans can only act out of their own self-interest). The obvious steps are taken to counter-evidence the (prima facie dubious) thesis: various acts of human kindness, sacrifice, compassion, and altruism, are cited. To this it is replied that these won't do; all of these, it is said, are simply cases of a more sophisticated, enlightened self-interest. The objector is forced, thus, to move on to more heroic, other-regarding actions, which are, in turn, explained away as "really" cases of egoistic behaviour, and so on, until even the most apparently obvious and grand cases of altruistic conduct have been tried, and found wanting. It is then clear that nothing, but nothing, can count against this egoism thesis, that it is actually an a priori assumption which will be used to interpret all data. The

anti-egoist therefore, simply cannot win and reflection on why this is, so reveals to us that it is something he shouldn't get especially disturbed about. The objector can at this point, turn the tables on the egoist and ask him where this a priori assumption comes from, and of course, he will be very unlikely to be able to tell us. The best he will surely be able to do is to cite cases of egoistic behaviour, but then the thesis itself notwithstanding, the objector can cite cases of other-regarding conduct, and the egoist is back to square one. (An excellent discussion of egoism and the unfalsifiable approach, occurs in John Hospers' Human Conduct.) Other examples of this technique spring to mind fairly readily. The area of political and ideological rhetoric is a notorious minefield of unfalsifiable arguments. No one who gives any serious attention to political debate can fail to be aware of both the left-wing ideologue who explains away all apparent counter-evidence to this claims, as examples of befuddled bourgeois conceptualising and his right-wing counterpart, who explains away all apparent opposition, as cases of mistaken proletarian conceptualising, born out of resentment, envy, distaste of success, or whatever. Some varieties of religious believer can hardly be forgotten either: there is the believer who appears to throw up some tenet or other of his belief, perhaps even God's existence itself, as a matter for debate, but explains away all apparent weaknesses in his position in terms of the belief itself. Indeed, there is also the believer who has the bottom line of "it's a matter of faith", i.e. the believer who is actually prepared to accept that there appear to be reasonable grounds for doubt regarding his belief, but still thinks that one can go on believing (whatever such a "belief" would actually be like) and, in fact, that it is a measure of the rock-solid strength of his belief that he can recognise the weakness of his evidence, and yet believe nevertheless. Within theology and the philosophy of religion itself, the problem of evil has also suffered at the hands of "explaining away" participants. If rational appraisal does not seem to afford a squaring of God's benevolence with God's omnipotence, then various re-definitions of good and evil can be attempted, or hitherto unsuspected shortcomings in humans' "way of seeing things" can be provided, disanalogies set up between "our

perception of good" and "God's (superior) apprehension of good", until it becomes clear that, no matter the evidence, God cannot lose. Psychoanalysis (some variants, at any rate) has also been discredited by this feature: any opposition on the part of the patient or critic to the details of the working of the Unconscious, as documented by the analyst, are explained away as cases of resistance to the (painful) evidence of the Unconscious. Like God, in the previous example, the Unconscious cannot lose. Finally, there exists a class of man or woman, who hold an unfalsifiable position to the effect that the opposite sex are intrinsically evil; no kind of conduct on the part of woman can save her or her sex from this kind of man, and no kind of conduct on the part of a man can save him or his sex from this kind of woman.

It should, by now, be clear enough what I'm talking about here, and it shouldn't be altogether inimaginable where it gets a hold in free will/determinism. In a section of the The Refutation of Determinism, "On Not Being Able to Help It: Power and Responsibility", Ayers considers the notion that the victim of brain-washing who "could not help" being converted and "cannot help" thinking the way he does, is no different from the rest of us. That is, is not everyone influenced by background, education, and pressures of society? For instance, few in Russia are pro-capitalist, few in America are pro-communist. In the light of my exposing of the phenomena of unfalsifiable arguments, we should immediately be on our guard concerning the claim that we are all victims of brain-washing. It appears to me, as it does to Ayers, that the claim is a spurious one and, as I hope to make clear, can only be sustained at the price of collapsing into a non-claim. Before mentioning anything else, I hope it is obvious, by now, how such an argument would proceed: there appear, at least, to be some examples here and there (to keep the claim modest) of people who are not altogether "brain-washed", and the propounder of the "universal brain-washing" thesis would attempt to explain these away, as cases of more insidious, undercover brain-washing, which leaves us all (himself included) absolutely no different from (say) the fire and brimstone religious believer. According to this, none of us ever

achieves any genuine liberation from, or mastery over, received ideas, ideologies, pressures, etc. Before concerning myself with anything said specifically by Ayers, it is, I think, more than worth noting another upshot of several of these unfalsifiable theses, including this one. And that is that they often have the effect of quite arbitrarily, needlessly, and pointlessly shifting the terms of the issue, whilst leaving what is important and substantive completely untouched. For instance, we could, if we wanted to placate the especially stubborn philosophical egoist, say that we are forced to accept his bleak thesis that all human actions are motivated from self-interest, whilst still preserving the distinction we regard as significant, i.e. we could then say simply that the significant distinction is between those actions which are apparently motivated from self-interest, and those actions which are apparently altruistic. We could grant the egoist that we are wrong to talk of a straightforward distinction between self-interested and altruistic behaviour, since we now know that all actions are really self-interested, but we simply shift the parameters of the issue, which now concerns a distinction between the "apparently selfish" and the "apparently altruistic". In all our subsequent talk concerning moral theory, praise and blame, culpability, etc., we simply remember to slip in the word "apparently" before talking of anything relating to selfishness and altruism, and the apparently selfish and apparently altruistic have exactly the same import as the selfish and altruistic respectively had before. So, for all the endeavours of the egoist, nothing has changed. A claim having a similar status, I think, is the one that has appeared in sceptical epistemologies from time to time, namely that it is possible that all life is a dream - this, I think, merely shifts the distinction from dream/veridical to seeming dream/seemingly veridical. I myself am not sure (very contentiously) that similar remarks don't apply to Kripke's attempts at doing away with contingent identities, though I don't wish to open the debate here. So, back to universal brain-washing.

We could, I think, appease the propounder by saying that we accept his claim, whilst it being the case that nothing substantive

has actually altered. People's behaviour, and the data with which we are presented daily, would not seem to be any obviously different, in the light of the acceptance of this claim and, à la the previous examples, we would learn to simply modify our talk, i.e. instead of distinguishing between those who are brain-washed and those who aren't, we would start to distinguish between those who are apparently brain-washed and those who apparently are not. The interesting questions would still be there, and the situation would not seem to have altered significantly from the one prior to our acceptance of the thesis of universal brain-washing, i.e. the position where we granted at least some people some independence of mind (let's call it) - the only difference is that we now slip in "apparent" before "independence of mind". Or we may want to talk of degrees of brain-washing: our most intellectually independent subject is simply the one who has been brain-washed least. "But so what?" one is tempted to reply "what does brain-washing then amount to?" It now seems an appropriate time to cite the examples which Ayers (quite reasonably, I think) regards as counter-evidencing the claim that we are all victims of brain-washing. Ayers notes that people can and do escape from the ways of thought and behaviour of their youth, recognise for themselves in the once-admired teacher a man of small prejudices, and diagnose and control in themselves tendencies they despise in their grandfathers. We can easily imagine the determinist coming back with the rejoinder that the nature and extent of these "liberating" processes are themselves determined by heredity, environmental factors, etc., i.e. the brain-washing is simply identified a little bit further along the line. And it should not be difficult to see the old spectre of non-testability/unfalsifiability looming large here. It seems to me to be a straightforward and indisputable fact that the kind of things we have just seen Ayers cite do, in fact occur. The determinist would be unlikely to dispute this, but would question what his opponent is claiming for them. That is, the anti-determinist (e.g. Ayers) is claiming that the occurence of the processes in question demonstrate that we are not all, in fact, victims of brain-washing, that we are not all creatures of heredity, upbringing and environment, and the character that these have

bestowed upon us, whilst the determinist is saying that they demonstrate no such thing. It should be obvious by now that the appropriate question to ask a determinist arguing like this is that of precisely what, if anything, could count against his claim that we are all helplessly imprisoned in an outlook we didn't even create: if he cannot cite anything which could do this job, then it becomes very difficult to see what his claim actually is. What would he be meaning when saying that despite appearances, no one actually achieves any real liberation from the sad brain-washing which they have suffered? The ordinary evidence, taken at face value, just doesn't seem to support this determinist claim, at all.

However, where I think Ayers is a bit slack here, where he doesn't, I think, do full justice to this (ultimately unfalsifiable) empirical determinism/incompatibilism, is that he doesn't seem to recognise the extreme lengths to which this "explaining away" technique in defence of the "universal brain-washing" thesis can be deployed: we can go on and on, until we cite one's most critical, rational and meticulous examination of received ideas (including, we could say, a scrupulous examination of the critical processes themselves), as evidence that one has gained some liberation from even the most insidious, secretive psychological manipulation. However, the "unbeatable" determinist may well reply that, even at the greatest possible level of critical awareness, we are still, whether or not we are aware, embracing (for instance) ideologies and rhetoric, not essentially of our own making, and consequently, we are never really anything other than brain-washed. Such ideologies may only be those as widely shared and apparently innocuous as those of physical objects, persons, private and public, myself and other, subjective and objective, etc. So (it goes on), it is not only the heavily indoctrinated Marxist, capitalist, Muslim or Christian or whatever, who is in the grip of a conceptual apparatus (say) which he is apparently unable to transcend, criticise, or detach himself from, but indeed, we are all similarly imprisoned. Indeed (it continues) philosophical discourse and technique, the practices of critical analysis, are themselves only another form of rhetoric, another "set of metaphors" (as is sometimes said), another

ideology. That (it will be said) is the very best which we can do: though we may avoid more primitive ideolgies, we may ultimately be in the grip of the rhetoric of criticism, of asking for reasons, evidence, etc. And (it may conclude) in devising this rhetoric, we were in the grip of yet another one, and even the most fundamental rhetoric we can reach does not appear to afford vindication from anything outside itself. All of this should ring loud bells in the ears of those acquainted with Post-Modern trends in philosophy, according to which philosophy is effectively out of a job.

It should be no surprise by now, however, that I believe that any appearance what has just gone may have of being especially radical, or of having fundamental consequences vis a vis our freedom, is an illusory one. To say that someone is (say) imprisoned in a rhetoric of critical analysis, of not accepting things without evidence, etc., is really to say nothing at all. So far as one stays on this straightforward, empirical level, it would seem that the greatest level of freedom which one can possibly have regarding one's environment, upbringing, etc. surely consists (in part, at least) of being in possession of, and applying to the attitudes, values, etc. in question, the most precise and rigorous critical techniques. If the determinist says that even this won't do as regards rescuing us from the spectre of brain-washing, then it becomes very difficult for us to see what could possibly do the job - it begins to look as though this particular determinist has no coherent notion of indeterminism. So, if it is claimed that, even as I write, I am in the grip of a rhetoric, i.e. the rhetoric of argument, debate, rational analysis, the rhetoric which comprises the stage on which even debates about ideology, rhetoric, free will/determinism and responsibility take place, then I hardly need lose any sleep over submitting to this claim, where the issue of my part in my actions is concerned. How would I possible be any more free with regard to the ideas I have received than to be "imprisoned" like this? It seems to me that a determinist going in for this line of attack is doing little more than making the (ultimately trivial, I think) Wittgensteinian point that we are, each of us, born into a world which is already theory-laden, where

even what appear to be our most fundamental perceptual judgements entail, and are a part of, an entire conceptual framework, which is not obviously without features of arbitrariness. It is what is alleged to follow from this fairly uncontroversial claim that I do not see. If we reach awareness of even this feature of our lives and constantly keep alive in our minds the possibility of alternative conceptual frameworks, then what can this determinist/incompatibilist possibly be wanting from us if he says that even this doesn't mean that we are not all victims of brain-washing? I feel it worth stressing again at this point, that the problem of free will is simply not to be settled in this fashion (if at all): it is not a matter which can be resolved by looking at simple, straightforward ordinary empirical data, or by recounting simple empirical power-ascription techniques. There may well be reasons (which may be discoverable at some point) for believing, à la Ginet's 'H' hypothesis, or Honderich's ('One') determinism, that we cannot give genuine content to the notion that we can act other than how we do act, but the kind of ordinary, empirical determinist/incompatibilist efforts are not among them. The truth or falsity of determinism or compatibilism is simply not discoverable as painlessly as the likes of Ayers, Lehrer, or the empirical determinist/incompatibilist seem to think (otherwise, again, it would surely not have had the history in philosophy which it has had - I would be unlikely to be writing a thesis on it just now); it is a matter which could only possibly be settled beyond the "veil of appearances", so to speak. Indeed, I make no apology for harping back to this point, to the significance of Van Inwagen's (M) hypothesis. I hope that the poverty of the "explaining away" attempt at establishing our impotence in the world is clear enough. This garden-variety determinist position can, I think, be exposed for what it is. However, it must be added that, should we have failed to be aware of it before, the logical posibility of (M) or something significantly like it being true, exposes also the very serious limitations of the kinds of reply made by someone like Ayers to the garden-variety determinist. What this kind of reply can do adequately is knock down the empirical determinist/incompatibilist. But then, as I hope I have made clear, one is punching at a man of

straw in so doing. What this kind of reply cannot do however, is counter-evidence those less simple-minded and more challenging determinist thesis: (M) or anything significantly like it, such as Ginet's H hypothesis, or Honderich's ('One') determinism. The fact, for instance, that one does diagnose and control tendencies one despises in one's grandfather or that one does realise that the once-admired teacher is a man of small prejudices, does, I think, straightforwardly refute the claim that one is necessarily doomed to be imprisoned in the ways of thought and behaviour of one's youth. However, what facts such as those do not counter-evidence is (M), or any hypothesis significantly like it. All of the phenomena which Ayers documents as supposedly counter-evidencing determinism may be real enough (and I've already said that, largely, I think they are), yet (M) or H, (for instance) may be true, with the fundamental consequence that we are not responsible for our actions. This empirical approach neither demonstrates that we can't do otherwise (despite the empirical determinist/incompatibilist), nor that (despite Lehrer and Ayers) we can do otherwise.

Indeed, we can now get to the bottom line. Much of what I have said already about the ultimate poverty of this empirical determinist/incompatibilist approach is itself no accident, is really the result of something going on at ground-floor level. We can get some inkling of this by noting what I feel is the final irony of this particular empirical approach. That is that a genuine analysis of the ordinary evidence firmly suggests neither the conclusion that we can't act other than we do act, nor its opposite, that we can do otherwise. That is, the ordinary evidence, it seems to me, suggests that we have a real problem on our hands, which is not going to go away or be settled easily. Our responses and practices, in fact, seem to me to betray a frequent lack of confidence on the questions of freedom and responsibility - we often seem to be unsure how responsible we ought to hold someone for their actions. We seem to often talk of degrees of freedom. People are, no doubt, presented with ideology, rhetoric, etc. with vastly varying levels of vigour, zeal and reinforcement. They also seem to

be given vastly varying facility for criticising it, and liberating themselves from it. It also seems sometimes that individuals have varying capacity for liberation e.g. there is the occasional Russian who is anti-communist, despite not having been presented with an obviously different rhetoric to the vast majority who are not anti-communist. So, in individual cases, we find ourselves asked to ascertain (possibly among other things) how vigorously a person has been presented with the given mind-set(s) relevant to doing a particular action, having a particular attitude, etc., what facility they can be reasonably thought to have had (or ought to have had) for detaching themselves from the mind-set(s) in question, the significance of particular contingencies of the individual situation, etc. Establishing even what are the mind-set(s), etc. which have been operative in the particular situation will not itself always be easy. So, it seems to me that <u>even the ordinary evidence</u> suggests that the whole business is problematic; like many issues, there may well be extreme cases about which we feel in no doubt, but a whole load of problematic ones in between. For instance I, with a critical training, would have little excuse, it seems, for entertaining attitudes which result from a dubious rhetoric I have been given the facility to see through; I cannot reasonably be excused, it seems, by putting forward such a rhetoric as a reason for (say) a morally questionable act. On the other hand, we may well excuse a person doing the same if the person has clearly been allowed little opportunity to transcend the rhetoric in question, and has had it presented to him in a very zealour manner. But the enormous number in between present no obvious answer, and each seem to demand a careful, precise appraisal. Far from the ordinary evidence suggesting free will, it seems to me that the daily kind of exercise I have just mentioned is another major source of evidence that we are aware of the reality of the worry of determinism. And it is when we attempt to account for this irony that it becomes apparent what has really been going on, at bottom, the whole time, and in terms of which so much else which has been said with regard to this particular area, begins to make sense. I talked earlier about the ground-floor reasons why the 'A' slots within Ginet's 'H' hypothesis could not be occupied by psychological

concepts. I explained this, really, as the outcome of Davidson's thesis (or the thesis attributed by Kim to Davidson) of Psychophysical Anomalism, according to which the psychological is simply not the kind of thing to enter into lawlike relations with the physical (supplemented by Psychological Anomalism, this amounts to the all-embracing thesis that the psychological is just not the kind of thing to feature in laws). I put this forward as the reason why contingent necessitation between the psychological and behaviour is not really possible, this in turn explaining why any set of relations which appeared to counter-evidence this had to be revealed, on inspection, to turn on "explaining away" kind of moves. Precisely the same kind of thing is going on here, in the case of empirical determinism/incompatibilism. This is really the reason for the final irony I have just documented and for much else I have said in this part. What is going on, at bottom, is this: the kind of factors put forward as evidence for our being unable to act other than how we do act are not the kind of thing to enter into lawlike relations with behaviour. That is really why this approach is ultimately empty, and I hope that, after all I've said already (not only with regard to empirical determinism/incompatibilism, but also with regard to H, contingent necessitation, and Davidson), it makes perfect sense that any supposed argument "evidencing" empirical determinism/incompatibilism must collapse into an "explained away" non-argument. I hope that it is entirely clear that the recurring conjunction of "explaining away" laws and vocabulary unsuited to lawlike relations is no accident. They are bound up with one another. For instance, the claim, cited by Ayers, which I mentioned a short while back, that we are all doomed to be prisoners of heredity, upbringing, environment, etc., cannot be true in any substantive sense: ordinary inspection seems to reveal its falsity (as Ayers points out), and it can only appear to be true by being explained away into a non-claim. The likes of social and environmental factors (just like the psychological) just seem to be ruled out, as regards entering into genuine lawlike relations with behaviour. As with the psychological, there always appear to be too many variables operative in any given case for lawlike relations to be any sort of possibility. For instance (to go back to another

example of Ayers'), some do recognise in the once-admired teacher a man of small prejudices. Others don't. There cannot be the slightest hope of accounting, in terms of laws, for why some still admire their teacher and others don't. As I've said, apart from all else, there are a myriad of variables to be considered. It is now banal, I hope, to say that, when we are talking about determinism, we are talking about laws, and I hope it is now fairly clear why the only kind of vocabulary suited to laws is the physical. I hope that I have done enough by now to demonstrate that laws must, qua laws, be physical. When we are talking about laws, we are talking about particles, speeds, neuron firings, nerve cells, etc., we are not talking about desires, hopes, wishes, fears, purposes, or heredity, environment, social pressures, intelligence, indoctrination. And, since determinism entails laws, when we are talking about determinism, we are talking only about particles, speeds, brain-states, never-cells, behaviour and the like. That is why care is required with Ginet's locution of "physiological-cum-environmental" laws. Despite the reference to environment, I hope it is clear by now that we cannot be speaking of any social or psychological factor here. Even "environmental" can only have strictly physical import. Whilst this is not the place for a discussion of the philosophy of the social sciences, this realisation about the appropriate (and naturally, also inappropriate) vocabulary for laws must surely have important consequences for the range and scope of the social sciences. I would imagine that a lot of the significant remarks made by Davidson about psychology in "Psychology as Philosophy" would be paralleled in other disciplines e.g sociology. I cannot really pursue this to any serious extent here however. What is important is that what I have called "empirical determinism/incompatibilism" must like "psychological" determinism, fail right at the outset because its vocabulary cannot support it. If determinism is true, this is something (as I've already said) which can only be known through knowing about particles, neuron firings, and the like. Of course, at this point, it should hardly need repeating that it is hardly agreed by everyone that the actuality (should it turn out to be the case) of nerve cells, neurons, etc. behaving in a deterministic

fashion, would entail that we can never act other than how we do act, and it is now time for me to consider another compatibilist tradition, an approach which has been accorded a degree of respect which persistently surprises me (as I said), this being conditionalism. Even Van Inwagen (who I am more often than not in agreement with on issues in free will) seems to me, whilst making numerous sound comments on it, to miss some fundamental shortcomings. Indeed, for reasons which I hope I will soon make clear, the whole issue of conditionalism seems to me to be, in fact, little better than an irrelevance with regard to questions of free will, determinism and compatibilism.

8. CONDITIONALISM

It is difficult not to feel the gut, intuitive appeal of conditionalism. I talked away at the beginning of the preverbal and precultural sources of a great many of the philosophical problems, how this feature discredits specific tendencies within recent philosophy, and how free will/determinism is but one example where this aspect should be appreciable. And indeed, once one has "felt" the free will problem as one of those kind of intuitive, half-formed anxieties, it is, I think, a less than enormous leap to "feel", in an equally kind of gut fashion, that conditionalism is the obvious way of securing compatibility between free will and determinism.

We have seen attempts at useful articulation of the thesis of determinism, in the shape of Ginet, Honderich and Van Inwagen, but no matter what precise articulation is eventually accepted as a successful conceptualisation of the problem, what is in no doubt is that what primarily generates a worry about determinism and free will (and what, it seems, any formalising of the problem must recognise), is the fact (or supposed fact) of causation. (Or, at any rate, as I've already suggested, the concern that the causation which operates specifically in the case of human action may entail determinism with regard to human action, leaving it difficult to see how free will could be true and how in turn, ascriptions of responsibility to persons could be warranted.) And what precisely makes conditionalism seem like a knight in shining armour for free will, is that, far from trying to argue away or marginalise the fact of causation, the conditionalist welcomes it with open arms and indeed, utilises it to supposedly reinforce our belief in our free will. I have not yet stated the conditionalist thesis itself and, before I do, I can, I think set the stage for it by posing what seems to be, really, a rhetorical question: how could an action be any more genuinely mine, my action, if it is caused by a choice, wish, fear, hope, or whatever ("intentional state" may do, I suppose) of mine? Despite the fact that it is, first and foremost, a theory about meaning, it is this possibility which conditionalism (if correct) seems to open up. My actions may indeed be as

straightforwardly caused as anything else which happens in the universe (so goes the conditionalist argument), but, unlike the other effects in the universe, what causes them are choices, etc. of mine. I decided to press the "coffee" button on the drinks machine, which caused my pressing of the coffee button, but then, had I decided instead to press the "tea" button, then my pressing of the tea button would have been caused. All I needed to do was choose differently, and a different effect would have issued. What could be a more obvious or a more harmonious way of squaring determinism with free will? That is, suppose (the argument goes) that the facts about human action and causation really do entail determinism with regard to human actions: what if the actions themselves are determined by the relevant choices, intentions, purposes and the like? Freedom back, safe and sound. Everybody happy. (It may well occur right away that such a suggestion is itself looking inconsistent with what I've already concluded concerning the kind of thing which could appear in deterministic laws i.e. that, due to Psychophysical Anomalism and Psychological Anomalism, intentional states cannot appear in laws of nature, cannot determine, even if they can cause. This realisation will be seen to have a more than passing significance by the end of this chapter.)

Despite the history of conditionalism within philosophy, the motivation which is at the bottom of it is really as I've described. And in fact, it is the exact nature of this motivation which leaves it looking surprising to me, not only that conditionalism has the formidable philosophical history which it does have, but that no one, it seems to me, has ever made properly explicit exactly why conditionalism is a hopeless compatibilist project (I must include Van Inwagen here). Let me clarify.

As I hope will become entirely clear, conditionalism is, really, a theory about meaning. This feature alone should, I think, make us immediately suspicious. "If at any given time, there is only one physically possible future, then how can any facts about meaning secure free will for me?" Such a question, I think, expresses an immediate and obvious worry, and one which I think

continually haunts the conditionalist effort. The question also points us towards the fundamental reason why conditionalism must fail. That is, it seems too often generally assumed (by both conditionalists and non-conditionalists alike) that the issue is an entirely linguistic one, that if we can just arrange one or two little things conveniently within the linguistic realm, we can then leap across, as if by magic, into compatibilism/incompatibilism, and win the game for compatibilism. However, it is precisely this kind of assumption which seems to me to be mistaken. It seems quite clear to me that it is a case of putting the cart before the horse to think that the philosopher of free will who is desperate for a compatibilist solution, just sits there doing nothing, waiting and hoping that the conditionalists in the linguistic laboratories will be able to cook up something for him. That is, the reasons why I think conditionalism must fail are not any local, linguistic ones; indeed the linguistic analysis only turns out as it does, only looks as hopeless as it does, precisely because of much more fundamental and pervasive features of the question of free will. It is no mere linguistic accident, which the would-be compatibilist can only curse his luck over, that Van Inwagen's attempt, for instance, at securing a conditional analysis of 'could have' seems doomed (as we shall see). On the contrary, Van Inwagen's attempts only themselves fail, because the shallow hypothesis of conditionalism has no way of coping with what (as I have already said) is so fundamental to free will, i.e. conditions of choosing, willing, and the like. As I've already indicated, Van Inwagen doesn't bring this out quite as explicitly as he might; he may also leave one feeling that it is some little linguistic irritation that prevents the conditionalist coming to the rescue of the compatibilist. However, from what I have just said, I hope that it will be clear when I look at some of Van Inwagen's remarks on conditionalism, why this is not so. In short, I hope it will become clear that the conditionalist hypothesis writes itself off, at the very outset as any sort of account of the 'could have' relevant to free will. I hope that, by the end of this chapter, these suggestions will be wholly clear and defensible.

Conditionalism is accepted, in one form or other, by the great majority of the present-day defenders of compatibilism. The argument rests on a theory about the meaning of ascriptions of ability. According to the theory, ascriptions of ability are really disguised conditionals. For instance, according to one version of conditionalism, what the proposition "Smith could have saved the drowning child" really means is "If Smith had chosen to save the drowning child, Smith would have saved the drowning child". (Van Inwagen's example.) One point, of course, on which conditionalists differ is that of the proper content for the antecedents of their conditionals. Where one conditionalist will say, "If Smith had chosen...", others will say "...had willed...", "...had decided...", "...had set himself...", or "...had tried...". As Van Inwagen notes, if an adequate conditional analysis of ascription of ability is possible, it will assign of them conditional paraphrases of rather more complicated forms than the form exemplified by the above paraphrase of the "drowning child" example. For example, it could hardly be true that (Van Inwagen's example) "Napoleon could have won at Waterloo" really means "If Napoleon had chosen to win at Waterloo, Napoleon would have won at Waterloo". It is therefore expedient to concentrate on situations in which relatively simple conditional analyses are adequate if any conditional analyses are e.g. cases involving abilities to perform acts that don't involve the execution of elaborate plans or demand special knowledge or skill. However, even then, serious difficulties remain. Lehrer has, in fact, drawn attention to this. Consider the proposition: "Smith could have eaten one of the red candies". This is not equivalent to "If Smith had chosen to eat one of the red candies then Smith would have eaten one of the red candies". For suppose that Smith is pathologically afraid of the sight of blood, and that the candies are the colour of blood. Then it may well be that Smith was unable to choose to eat one of the red candies. And, in that case, he could not have eaten one of the red candies. Nevertheless, we may suppose, if he had chosen to eat one of the candies, he would have. As I've said, this kind of realisation is really, indeed, what will be seen (I hope) to be at the bottom of my fundamental dismissiveness towards conditionalism as an argument for

compatibilism i.e. the realisation, which is accepted by most garden-variety determinists even, in some form or other, and which I hope has emerged by now, that conditions of choice (and the like) are of monumental relevance to the question of free will. The question of free will loops back to take choice itself, and its conditions, into its purview - it cannot stop at the discussion merely of the path from choice to object. Of course, as Van Inwagen notes, efforts can be made to rescue Lehrer's example. We could try adding "...and Smith could have chosen to eat one of the red candies". But, of course, what would the second 'could have' mean? If it means the same as the first, and if the expanded conditional proposition really is equivalent to the proposition that Smith could have eaten one of the red candies, then the latter is equivalent to the statement

> If Smith had chosen to eat one of the red candies, then Smith would have eaten one of the red candies, and if Smith had chosen to choose to eat one of the red candies, then Smith would have chosen to eat one of the red candies, and Smith could have chosen to choose to eat one of the red candies.

This effort, however, contains the clause "if Smith had chosen to choose to eat one of the red candies" as well as an unreduced "could have", like its predecessor. But what is it to choose to choose something? One can, of course, as Van Inwagen remarks, choose to choose between two or more things: in choosing to drink wine, I may in effect be choosing to choose between drinking claret and drinking burgundy (Van Inwagen's example). But that is not to choose to choose something; what would count as a case of choosing to choose claret? It seems that the idea that the second 'could have' means the same as the original, leads to incoherency. Van Inwagen floats the possibility of the operation of another sense of "could have", but as we shall see, such a hope also seems futile. He suggests that we use 'COULD HAVE' to denote the required "new" sense of 'could have' and use 'could have' only in its power-or-ability sense. The conditional analysis then becomes:

x could have done y = if x had chosen to do y, x would have done y, and x COULD HAVE chosen to do y.

There would seem to be a condition that must be met by any adequate account of the meaning of 'COULD HAVE': 'x COULD HAVE chosen to do y' must entail 'x could have chosen to do y'. If this condition isn't met, then from the kind of example already considered, it should be easy enough to see that a counter-example to this definition can be produced. All we need do is simply pick one of those logically possible cases in which someone COULD HAVE chosen to do something, but could not have chosen to do it (and, of course, construct the example in such a way that his choice would have been effective). This will be a case in which our imaginary person could not have performed a certain act, though, according to the proposed definition, he could have. Consider again, Lehrer's example. If 'COULD HAVE' does not entail 'could have', then it may be the case, according to the proposed definition, both that Smith could have eaten one of the red candies, and that Smith could not have chosen to eat one of the red candies. This, of course, in so far as we are concerned with powers or abilities, is a nonsense. However, like Van Inwagen, I do not myself know any way to define 'COULD HAVE' that will meet this essential condition. And at any rate, as Van Inwagen points out, if this were possible, we would surely be able to give 'COULD HAVE' a sense such that 'COULD HAVE' tout court entails 'could have' tout court, and what then would be the necessity for the conditional that is the first conjunct of the definiens of the conditional analysis?

There is (I expect it would occur), an obvious kind of manoeuvre (which Van Inwagen recognises and discusses) open to the conditionalist. We can note that all attempted, conditional analyses presented so far featured the agent's doing something antecedent to his performing the action in question e.g. choosing. What, however, if a conditional analysis were to be attempted which did not include the agent's doing anything prior to the action? (In fact, when I come to consider the notion of agent-causation, what I will then be considering is the idea of an agent's causing

something without first doing something else.) If we could do away with acts of choice and the like, we may be able to avoid the infinite regress we have seen generated (Ryle, of course, in The Concept of Mind[1], displays awareness of the worries of the infinite regress.) Van Inwagen explores what may happen if a conditionalist were to offer an analysis of 'could have' of this form:

x could have done y = if Rxy, then x would have done y, and A

where 'Rxy' represents some condition on x and y such that it is possible that this condition hold without x's doing anything antecedent to his doing y. 'A' may be any sort of qualification. Such a format would allow us to stave off objections such as the ones I have already cited e.g. consider "if x had wanted to do y (more than anything else), then x would have done y." (Of course, such an analysis would take on board problems which I cannot tackle here, concerning the purported relationship between motivational strength and action. Irving Thalberg has written an excellent essay on this topic entitled "Questions About Motivational Strength"[2].) Wanting to perform an act (unlike choosing to perform an act) is clearly not itself an act. Imagine "If he had wanted to eat a red candy then he would have eaten one". According to this approach, it would be no genuine objection to ask "But could he have wanted a red candy?" This is because 'could haves' only get off the ground in the first place with the existence of given wants. As Van Inwagen notes, (because 'could haves' only enter the fray after the presence of wants is given), to attempt this kind of objection, to ask a 'could have' about a want itself, would be to make a kind of category mistake, would be to introduce some unspecified 'possible'. As we have seen, according to the 'choice' analysis "it was not within his power to choose to eat a red candy" seems to imply that it was not within his power to eat a red candy; however, according to this new analysis, its not being possible that he should have wanted to eat a red candy does not have this implication, since (as I've said) 'could haves' only become an issue once wants are given, it makes no sense to ask a 'could have' question about a want itself. This feature is really, in fact, the

downfall of the analysis in question. Van Inwagen suggests that we look again at the "red candy" case. Smith doesn't want to eat a red candy. The very sight of them, we may suppose makes him feel uneasy. But let us suppose for the moment that he did want to eat one of the candies. Presumably he could. But also, presumably, he would not be afflicted with his neurosis which, in actuality, renders him unable to eat one of the candies. So it would seem that both the statements "If Smith had wanted to eat a red candy, he would have" and "Smith could not have eaten a red candy" are true. That is, assuming that a coherent, meaningful sense can be given to the second statement in the face of the truth of the first one, which, of course, it can be. So the attempted new analysis must fail. We could, of course, attempt some qualification, some further condition to be inserted in the space occupied by 'A' in the abstract schema. Lehrer, for instance, makes a suggestion which could be incorporated into the analysis:

> x could have done y = if x had wanted to do y, x would have done y and it is false that if x had wanted to do y, x would have possessed some advantage with respect to doing y that x did not actually possess.

As Van Inwagen says, this would not have the consequence that Smith could have eaten a red candy. For, though it is true that Smith would have eaten a red candy if he had wanted to, it is also true that if he had wanted to eat a red candy, he would have possessed an advantage with respect to eating a red candy that he did not in fact possess: freedom from his neurosis. However, this effort is also pretty hopeless. Van Inwagen provides the example of his wanting to fly to Washington: if he wanted to fly to Washington, he should come to possess a certain advantage with respect to flying to Washington: a reserved seat on a Washington bound flight; and this advantage is one he does not actually possess. But this fact obviously does not entail that he could not fly to Washington. In order to cope with this difficulty, it is clear (that is, if there is any way of coping with it) that we must involve ourselves in a minefield which I think it unprofitable to tread any further into

(and I hope the defensibility of stopping here will become clear).

Van Inwagen, around this point, adopts what I think is an extremely useful and shrewd tactic. He supposes that we have found (whatever it may be) the best possible conditional analysis of ability, and he calls it simply the Analysis. The Analysis is right if any conditional analysis is right, and wrong only if no conditional analysis is right. If the Analysis is wrong, the conditionalist may as well pack up and go home - he can do nothing to help establish compatibilism. Van Inwagen notes that, so far as supporting compatibilism is concerned, the Analysis does very little for us, unless we have some reason to think it is correct. That is, some or other sense of 'could have' may afford a conditional analysis but, unless the 'could have' relevant to free will likewise affords one (which would be the truth of the Analysis), then conditionalism can do nothing for compatibilism. As I've said, however, it is no accident that the Analysis is false. The main thrust of my diagnosis is something which I have already given an outline of, and which I will go into some detail about in a moment, namely that (as should be quite clear by now) free will is contingent upon, not only the leap from choice or wish to action, but also the genesis of choice or wish itself. This realisation finds expression through several contributors and indeed, from much of what Van Inwagen says in the rest of his Essay on Free Will, the respect he accords conditionalism as a compatibilist thesis surprises me. Watson, for instance, in his essay "Free Agency"[3], distinguishes between freedom of action and freedom of will, the former turning upon the leap from choice or wish to action, and the latter concerning itself with the conditions of choices or wishes themselves. It seems to me, that conditionalism can only essentially concern itself with the former. There can be no doubt by now however, that free will takes the latter into its purview also. Anthony Kenny, in his essay "Freedom, Spontaneity and Indifference"[4], utilises the roughly equivalent Humean distinction between liberty and spontaneity and liberty of indifference; conditionalism, by its nature, can only take the former into its purview, whilst the issue of free will extends into the latter.

Similarly, whilst I will conclude that the efforts in question are ultimately a pretty futile contribution to the free will/determinism issue, it will be clear when I come to consider those efforts shortly, that Harry G. Frankfurt is also more than aware of the questions freedom can ask of choices and the like themselves. (Whilst I don't have time to pursue it here, there are, of course, parallel issues of a crucial nature in social and political philosophy e.g. what kind of desires and choices should simply be satisfied and what kind should we endeavour to change?) And as I have also just documented, even the strictly empirical or "explaining away" propounder of determinism/incompatibilism begins his assault by focusing on conditions of choices and wants, and the significance of this fact for what I am saying here is in no way undermined either by the fact that the evidence does not support his conclusions, or the fact that the kind of thing he puts forward as evidence is not (due to Psychological Anomalism and Psychophysical Anomalism), the kind of thing which could evidence determinism. Waiving for the moment the rights and wrongs of empirical determinism/incompatibilism, or those of the efforts of Frankfurt, Watson or Kenny, the central point is by now, I hope, clear enough. I can become much more specific about it.

I have already (in the preceding discussion of Lehrer's compatibilism) mentioned Van Inwagen's fanciful (M) hypothesis. It seems to me quite consistent for there to be some or other conditional sense of 'could have', and (M) to be true. That is, there could be a conditional analysis of some or other sense of 'could have', and yet no one be able to act other than how they do act. Let's say that I decide to have an orange juice, but also that had I chosen instead to have a lemonade, then I would have had a lemonade. Let us also imagine a fairly simple conditional analysis, i.e. let us say that

> x could have had a lemonade = if x had chosen to have a lemonade, and A, then x would have had a lemonade, where 'A' is some unspecified set of conditions (indeed, it probably wouldn't do any harm for what I want to say if we imagined A to

be absent from the antecedent of the conditional; my suspicion is really that the simpler the conditional is here, the better).

So according to this, if I had chosen to have a lemonade, and A, then I would have had a lemonade. But since my decisions are caused by a device implanted in my brain by Martians, a device which caused me to opt for an orange juice on this occasion, I could not have had a lemonade. In the only sense which matters, the situation here is exactly the same as it would be if events in the past conjoined with the laws of nature (neither of which are up to me) were sufficient for my having an orange juice, and not a lemonade. Specifically, I have access, in such a case, to only <u>one</u> possible world, namely the actual one, in which I have an orange juice and not a lemonade. Quite regardless of whether my choosing to have a lemonade would go even some of the way to facilitating my having a lemonade, the brute fact is that I have no access to a possible world in which I choose to have a lemonade. That is, I am not free to satisfy one of the conjuncts of the antecedent. This feature is really the fundamental poverty of conditionalism in so far as it announces itself as an argument for compatibilism i.e. the failure to even address the question of access to possible worlds, of whether I ever have access to a world in which I choose other than how I do, in fact, choose. There may have been nothing stopping me from having a lemonade if I had chosen to have one (I had freedom of <u>action</u>, or liberty of <u>spontaneity</u>), but I could not have had a lemonade (as a result of the Martian device, freedom of <u>will</u>, or liberty of <u>indifference</u>, is lacking). Conditionalism just does not seem to be the <u>kind</u> of thesis which could square free will with determinism. I may be accused of simply re-stating the kind of thing which Van Inwagen said with regard to the "Smith and red candy" example, of having us back at square one, but I would reply that the conditional analysis (where it is used to support compatibilism) <u>takes</u> us indeed, straight back to square one.

I hope I can make this even more clear by cashing it out where I think it really matters. I have already said that, whilst there are, obviously, important differences between hypothesis (M), and

something like Ginet's H hypothesis or Honderich's ('One') determinism, (as Van Inwagen recognises), the parallels are also more than useful. And I now want to try to show that, even if conditionalism has any currency as an account of any 'could have', free will is no more safeguarded in the light of H or Honderich's determinism being true, then it would be in the light of (M)'s being true. If H is true, then everything I do is contingently necessitated by physiological-cum-environmental laws. My taking orange juice instead of lemonade, due to physiological-cum-environmental law, could not have happened otherwise. It may be the case that had I opted for lemonade, I would have had lemonade, but what I will opt for is itself contingently necessitated. Indeed, the laws which contingently necessitate my taking orange juice have no place whatever for my choices or wants; all they care about are what (literally) goes into my mouth. Whether they are true is another thing again, but if they are true then (as the likes of Van Inwagen's two Formal Arguments for Incompatibilism mentioned up till now suggest) it seems very difficult to see how we can have any free will, to see how we can do other than what we do.

To reinforce this point, it may be of more benefit to look at Honderich, because Honderich seems to take slightly more seriously (or deal rather more cautiously with) the psychological; choices, wishes and the like which are so central to the Analysis. Ginet seems to merely make it his business to explain why psychological properties are not the kind of thing which can appear in laws, without making much of an effort to explain the impression of psychological efficacy, i.e. how it could square with H. Honderich, on the other hand, does do something toward the latter end, in the shape of his Correlation thesis. It may be wise to re-state the Correlation thesis at this point:

> Any particular description of consciousness D is true of an individual if and only if his brain or part of it is in one particular state or sequence of states S.

Remind ourselves of Honderich's other two premises, that states of

the brain are effects, the effects of other physical states, and that states of the brain are causes, both of other states of the brain and also of certain movements of one's body (actions) and my use of Honderich should not be too difficult to anticipate. Again, consider some conditionalist analysis which (as before) includes my choosing as one of the conjuncts of the antecedent. I took an orange juice, but, had I chosen a lemonade, and 'A', I would have had a lemonade. My choosing an orange juice and not a lemonade, was, however, correlated with a brain state or sequence of brain states, which was itself the effect of other physical states, which weren't up to me. So I could not have chosen to have a lemondade; my choice could not have been different from what it was, any more than if (M) was true. When we focus on the genesis of choices or wants in this way, their conditions, should their conditions be the Martian device, "physiological-cum-environmental laws", straightforward brain states, or indeed maybe even something else altogether, we see the poverty and in fact (in my view) the ultimate irrelevance of conditionalism to the cause of compatibilism.

Not only this, but I hope that it has emerged by now, how it can be that conditionalism may give the appearance of establishing compatibilism. That is because, as a theory about meaning, it looks very plausible as regards freedom of action (or liberty of spontaneity). To say that I have freedom of action with regard to (say) having the lemonade which I did not choose is surely just to say that, had I chosen the lemonade, nothing would have impeded my path to it. Such a sense of 'could have' is undoubtedly a perfectly genuine, meaningful one. Questions about freedom of action are perfectly real questions. However the compatibilist illusion here results from concluding that 'could haves' pertaining to freedom of action are the only genuine 'could haves'. As I hope is clear by now, this seems wholly untrue to me. The 'could have' pertaining to freedom of will, liberty of indifference, access to possible worlds, seems to me a perfectly real one also. Not only that, but the latter seems to be the one which free will/determinism addresses, and which conditionalism seems unable to provide a proper account of. Whilst conditionalism may yet be correct as an account of the

meaning of the 'could have' of freedom of action, there seems no hope of its accounting for the 'could have' relevant to free will/determinism.

It still seems at this stage clear to me that should H or Honderich's determinism (or anything significantly like it) be true, then we cannot do other than what we do, we do not have free will (this seems to me as unambiguously the case as it would be if (M) were true). I repeat that conditionalism doesn't seem to be the kind of thing which could either counter-evidence (M), H or Honderich's determinism, or protect anybody from what appears to be the clear consequence of them all, namely that no one can do other than what they do. Likewise, with the consideration Van Inwagen makes of examining how the Analysis stands in relation to his First Formal (incompatibilist) Argument, Van Inwagen comments:

> The First Formal Argument is valid. Therefore, if the Analysis is correct, at least one of the premises of the First Formal Argument is false. (I <u>assume</u> this conditional is true. If it isn't, of course, then free will and determinism are incompatible <u>even if</u> the Analysis is correct.)

My own response to this shouldn't be difficult to anticipate. Van Inwagen's conditional is true, but, due to, really, the most basic features about free will, the Analysis is false.

Before ending this part, I think it worth pointing something out. This concerns what we could conceivably add to a fairly typical conditional analysis (e.g. 'I could have' = 'I would have, if I had chosen'), which would make it look like a case for the existence of free will. We can, I think, be given a hint towards realising what this is by noting something which is, I think, itself revealing about the status of conditionalism. This is that Ginet and Honderich, in their respective offerings, just don't discuss conditionalism. It doesn't rate a mention with them. They clearly don't regard conditionalism as any possible way of securing our ability to act otherwise, should the determinism they respectively

speak of be true. Similarly, in "The Conceivability of Mechanism", Malcolm doesn't mention it. Indeed, the comments of Malcolm in particular, which I have cited, should hint strongly at what would be needed to supplement conditionalism on behalf of free will. Remember that the antecedent features choosing, wishing, or some such intentional concept. So, I hope it isn't too difficult by now to see that what is really needed to construct an argument for the existence of free will, is an additional premise to the effect that the intentional enjoys (as well as the irreducibility, which Davidson speaks of) a particularly far-reaching, "untouchable" kind of autonomy, i.e. that the intentional does not refer us back to some more basic (to use Malcolm's expression) realm or vocabulary, which could explain exhaustively, be it the neurophysiological "physiological-cum-environmental" laws, or any kind of "laws of nature". As Malcolm has so precisely pointed out, if (for instance), basic neurophysiological laws are at the bottom of all our choosings, wishings, etc., and can therefore account for everything, entirely in the absence of invoking intentional concepts, then there is a perfectly clear sense in which my choosings and the like, don't really matter, in which they are superfluous to an explanation of my action. In Malcolm's aforementioned example, we could account entirely for the man's climbing the ladder purely by recounting neurophysiological laws, we needn't cite the desire to get his hat from the roof at all. His desire to get his hat from the roof refers us back to a more basic neurophysiological, which can exhaustively account for his climbing the ladder. Why bother with secondary, ultimately superfluous, concepts such as choosing, wishing, and the like, when we've got the basic tools to account for everything? It is, I hope, clear that if determinism is true, then conditionalism, through featuring intentional concepts such as choosing, is going to refer us back to something more basic, which is both not up to me, but accounts for (in terms of laws of nature) and therefore relegates (in an explanatory sense) my choosing, or whatever. If, say, Honderich's determinism is true, then choosing and desiring and any action which follows, can all be wholly accounted for by a more basic neurophysiological, behaving in accordance with laws of nature.

Assuming that Ginet is postulating a pretty similar sort of story, the truth of H would mean that my choosings and the like and any action which follows, can all be subsumed under more basic physiological-cum-environmental laws. It should be obvious why Honderich, Ginet and Malcolm, would regard conditionalism as something of an irrelevance, and I hope that my surprise at the respect afforded it by Van Inwagen is now looking reasonable. I hope that it is also entirely clear why I think that, in order for conditionalism to be involved at all in arguments in favour of free will, the conditional must take the form:

> x could have done y = Conditional statement, and Z
> where Z is some condition to the effect that the intentional cannot be subsumed under any more basic laws.

However, if it has remained rather hidden up till now, this "new" conditionalism may have brought something else to light. Right from the start of my look at conditionalism, it may have seemed that especially in the light of what has gone before, there was something slightly odd going on. Remember that conditionalism advertises itself as an argument for compatibilism,i.e. it is given that determinism is true, and the problem is to square this with our free will. Enter the psychological, our choices, wishes and the like as the free will/determinism hybrid. Therefore, it may seem reasonable to call conditionalism a <u>psychological</u> determinism. However, if what we have seen said by Davidson (or, at any rate, Kim's re-stating of Davidson), Ginet and Van Inwagen is correct, then there really can be no such thing as psychological determinism. As we have seen Van Inwagen, in particular, make admirably clear, when we talk about <u>determinism</u>, as opposed to mere causation, we are talking about <u>laws</u> and, as Davidson, Kim and Ginet all help clarify, the psychological is not the kind of thing which can feature in laws, which can have nomological links with anything else, be it other aspects of the psychological, or the physical (Kim's Psychological Anomalism and Psychophysical Anomalism). Davidson, in order to sustain a "psychological" determinism, is forced to "descend" to the more basic physical for his laws, with

the killing consequences for free will which I've already made clear. Honderich's position seems slightly different; in his Correlation thesis, he seems to have the (mistaken, I think) belief that there can be psychophysical nomological links, but this is not really what is important, however; what is important is that, like Davidson or Ginet, he believes the basic, subsuming laws to be discoverable at the physical level, with again killing consequences for free will I hardly need spell out again, and which Honderich fully accepts. So, it seems to me that there are two possibilities for conditionalism:

(a) if we do not add in premise Z, and determinism is true, then since (as we've seen), the only determinism it makes sense to speak of has laws discoverable at the physical level, then conditionalism provides no support for free will;

(b) we add in premise Z, in which case because of the aforementioned Anomalism of the Mental, adds up in fact to a denial of determinism. Z, by nature, precludes subsumption of the mental under more basic (physical) laws, but if determinism is true, there must be laws. But the psychological is not the kind of thing to feature in laws. Therefore, if Z is true, determinism is false.

That is, conditionalism can only be a part of an argument for free will if determinism is false. And this amounts to the claim, of course, that despite its firmly entrenched position in philosophical history as an argument for compatibilism, it can only do anything at all for the cause of free will on pain of also entailing incompatibilism. Conditionalism, at the end of the day, does no more for the cause of compatibilism, than do the efforts of Lehrer I considered a bit earlier. Another cheap trick.

As I've already said, the work of Harry G. Frankfurt in "Freedom of the Will and the Concept of a Person" can appear very much like another contribution to compatibilism. Since Stevenson, philosophers have been familiar with the term "Persuasive

definition"; there is also such a thing, though, as an evasive definition, and I will be wanting to claim that Frankfurt's attempt at disparaging the significance of determinism vis à vis free will, can only appear to succeed because his definition of free will is ultimately, an evasive one. It is another cheap trick.

Notes

1. In The Concept of Mind, Ch.III, P67

2. In Ernest Lepore & Brian McLaughlin, Actions & Events: Perspectivbes of the Philosophy of Donald Davidson.

3 In Free Will, ed. Watson.

4. In Essays on Freedom of Action, ed. Honderich.

9. FRANKFURT AND EVASIVE DEFINITION

> There is no more than an innocuous appearance of paradox in the proposition that it is determined, ineluctably and by forces beyond their control, that certain people have free wills and that others do not.
>
> - "Freedom of the Will and the Concept of a Person"

This claim occurs in the penultimate paragraph of Frankfurt's essay, and has a clear compatibilist look about it. As I've already suggested though, I think that the claim can only be sustained on pain of a re-definition (evasive re-definition) of free will, and one which is ultimately irrelevant to the substantive questions within free will/determinism. Frankfurt himself admits that the possession of this "free will" does not have the connection with responsibility which free will is usually taken to have; one of the major worries surrounding determinism of course (especially where it is conjoined with incompatibilism) is that free will is generally thought to be a condition of moral responsibility, but with Frankfurt's conception of free will here, whilst free will may be necessary for responsibility, it is (as he is aware) certainly not sufficient. Indeed, an examination of what Frankfurt means by "free will" will reveal, I think, that his position is not, in any significant sense at all, a compatibilist one. Despite the seemingly compatibilist paragraph above, his essay actually contains no contribution at all to questions of determinism, libertarianism, compatibilism and incompatibilism. The fact that the "paradox" he speaks of above is only an innocuous appearance is actually of no real significance, since "free will" is evasively defined, with the consequence being that free will and determinism look as difficult to square as ever before, and moral responsibility looking as out of tune with determinism as before.

Frankfurt is, in a sense, I think, hijacking the notion of free will, and I think that it is important to expose this. So what is he actually talking about? From the same essay:

> Besides wanting and choosing and being moved to do this or that, men may also want to have (or not to have) certain desires and motives. They are capable of wanting to be different, in their preferences and purposes, from what they are...No animal other than man...appears to have the capacity for reflective self-evaluation that is manifested in the formation of second-order desires.

To have a second-order desire is to want simply to have a certain desire; however, to have a second-desire volition is to want, not only to have a certain desire, but to want a certain desire to be one's will, i.e. to want a certain desire to be the desire which moves one to act. Frankfurt distinguishes the two by means of the following example: a physician engaged in psychotherapy with narcotics addicts believes that his ability to help his patients would be enhanced if he understood better what it is like for them to desire the drug to which they are addicted. Suppose that he is led in this way to want to have a desire for the drug. However, it is entirely possible that he does not want this desire to be effective. He may not want it to move him all the way to action, and indeed, he may prudently arrange to make it impossible for him to satisfy the desire he would have if his desire to want the drug should, in time, be satisfied. The physician, in this example, has a second-order desire to take the drug, but his second-order volition is not in tune with this. And now the punch-line - this is what really seems to be important to Frankfurt: according to Frankfurt, it is only because a person has volitions of the second-order (unlike animals and 'wanton' humans, as he puts it) that he is capable both of enjoying and of lacking freedom of the will. The possibility of second-order volitions is (for Frankfurt) a condition of the problem of freedom of the will:

> When we ask whether a person's will is free we are not asking whether he is in a position to translate his first-order desires into actions. That is the question of whether he is free to do as he pleases...Rather, it concerns his desires themselves (emphasis mine). But what question about them is it?

I think Frankfurt is completely correct here to suggest that the notion of free will, despite an entire tradition within compatibilism which I've devoted some attention to, loops back to embrace desires themselves within its purview. However, the question that Frankfurt thinks free will asks with regard to desires is simply not the kind of question I think it asks e.g. is the neurophysiological more basic than the (mentalistic concept of) desire, as (say) Honderich's three determinist premises document well? Can we give any sort of sense to the notion that the desires could have been any different than they were at the time in question (and if Ginet's 'H' hypothesis or Honderich's determinism is true, it makes it very difficult to see how this can be done)? Frankfurt has, I think, some curious things to say, on the questions free will asks of desires and, rather than criticise specific, individual suggestions one by one, I think, it more fruitful to cite a number of extracts from Frankfurt's essay, thus establishing an overall picture, which I can then go on to criticise:

> It is in securing the conformity of his will to his second-order volitions, that a person exercises freedom of the will...a person may have, especially if his second-order desires are in conflict, desires and volitions of a higher order than the second. There is no theoretical limit to the length of the series of desires of higher and higher orders; nothing except common sense and, perhaps, a saving fatigue prevents an individual from obsessively refusing to identify himself with any of his desires until he forms a desire of the next higher order...It is possible, however, to terminate such a series of acts without cutting it off arbitrarily. When a person identifies himself <u>decisively</u> with one of his first-order desires, this commitment 'resounds' throughout the potentially endless array of higher orders.

I am not happy with a lot going on here. For what it's worth (which I think will ultimately turn out to be fairly little), the claim that it is simply in securing the conformity of his will to his second-order volitions, that a person exercises freedom of the will,

appears to me, not only to be misguided, but odd in that Frankfurt goes on to recognise the possibility of desires and volitions of a higher order than the second. I am also not happy with the claim that there is no theoretical limit to the length of the series of desires of higher and higher orders, and that only common sense and economy of energy ever allow soneone to actually act. I am also doubtful of Frankfurt's suggested reason why the ending of the series of critical acts is not arbitrary.

Even within the whole Frankfurt framework, of marginalising the question of determinism with regard to free will (to which, of course, I am unsympathetic), it should not, I think, be especially difficult to see why it is entirely inadequate to say that one's freedom of will be exercised merely by the conformity of one's will to one's second-order volitions. This is because here is no reason whatever why second-order volitions cannot themselves come under the purview of freedom. I have already talked of brain-washing, ideology, rhetoric, etc., and whilst I've made clear what I regard as the ultimate significance within free will/determinism of an empirical examination of these phenomena, if we are to bear with Frankfurt and stick within his framework, it will, I think, very frequently, be one's second-order volitions, more than one's first-order desires or actions, which reveal the extent of these kind of manipulative processes. Indeed, it seems to me that, as Frankfurt has given them, the actual task of brain-washing and psychological manipulation, precisely is to create a particular second-order volition, though it may also be necessary to create a new first-order desire as well in some cases, i.e. the typical "manipulator" attempts to convince his subject that it is one particular desire, rather than another, which ought to move him to act, and if he succeeds in this, then he has created a whole new second-order volition and (as I've indicated) this creation of the new second-order volition may actually entail the creation of a new first-order desire as well which, of course, is the one he wants to move the subject to act. Indeed, it seems to me that it tends to be at the level of second-order volitions that we discover the operation of the likes of values, principles, theories of conduct

and of the good life etc., and the manner in which these have been acquired can be every bit as good a reason for regarding the agent as unfree to a significant extent, as can mere failure of his actions to conform to them. The results of manipulation, ideology, rhetoric, etc. can indeed reveal themselves straightforwardly in one's first-order desires and be opposed by a second-order volition which the agent has (critically) highly defensible reasons for holding, but just as often, I suspect, first-order desires will or will not be endorsed by a second-order volition which is, quite transparently, a consequence of manipulation and rhetoric which has never been subjected by the agent to any sort of critical appraisal. Consider the case of the man who has enjoyed a weekend drink for most of his life, and has never until now, pondered especially on whether he ought to do so. Then, quite arbitrarily, it seems he remembers someone having told him, at a very young age, that alcohol is wicked and that consumption results in eternal damnation for the drinker. For fear that this may be true, the desire not to drink then becomes among his first-order desires, and also becomes his second-order volition. Let us also suppose that, within a week or two, he has succeeded entirely, i.e. he has brought his actions into line with his second-order volitions, he has stopped drinking. In such a case, the subject has satisfied the condition propounded by Frankfurt for freedom of will, but even within this empirical framework, we would surely be very reluctant to say that such a subject is acting as freely as he could be. As I've said, consciously held principles and the like often enter the fray at this level, and our subject in this example could well turn his anxieties into a statable principle (e.g. "Carnal indulgence for its own sake is a bad thing"); here, the actions are not only in line with a second-order volition, but actually in tune with principles, which it may require some strength of character to practise. I hope it is clear, however, that the issue surrounds the principles, etc. themselves: in our example, the agent has not subjected his anti-pleasure principle itself to the slightest scrutiny. He has not questioned why he holds it in the first place, or wondered whether there are good reasons for holding it. It is simply a <u>given</u> for him; it may have been presented to him in

particularly expedient conditions with a degree of zeal and reinforcement which left him completely imprisoned within it, entirely unable to transcend it and criticise it. Quite regardless of whether such a detached appraisal would reveal the principle to be sound or otherwise, the fact that it is a possibility at all, which is not open to our agent, surely means we can say straight off that our agent could be more free than he in fact is. And, as I've suggested, I expect that a great many of the principles and values which can be found at the bedrock of people's "second-order" lives are taken on board in this uncriticised, untouched fashion e.g. love of God, love of country, professional success, public esteem, anti-hedonism, hedonism, pursuit of truth, various loyalties (I say nothing, of course, about whether there actually are grounds for defending any of these second-order motivations!).

From parts of Frankfurt's essay I have quoted, it seems that Frankfurt is dimly aware of something like this problem. However, both what seems to be his attempt at a diagnosis and his suggested solution, strike me as completely misguided. I believe, in fact, that in so far as one speaks within Frankfurt's parameters, both the problem and the solution are not so much a case of desires of higher and higher orders, but more a case of what is characterised by Nagel, as <u>standpoints of higher levels of objectivity</u>. In fact, this whole notion will become of crucial importance later, when I look at Nagel's sharp response to Strawson. Let me expound.

Frankfurt suggests that there may be a conflict of one's second-order desires, and that in such a case, a third-order desire may have to come in to adjudicate. There are a number of things I don't like about this suggestion, however. Generally, I am not happy with the concept of desire being allowed to be so broad that it extends from the most primitive (e.g. the wanton addict's desire for the drug) to whatever Frankfurt supposes goes on at the third-order level (and of course, possibly infinitely beyond, according to Frankfurt). I would prefer (for reasons which should shortly be made clear), to reserve and restrict the notion of a desire for the fairly primitive levels of the supposed hierarchy -

the wanting level and the wanting to want (perhaps) but really no more. Apart from all else, I don't really see any reason to believe that a person's motivational system could be as grand and complex as this. Perhaps I am just a miserably primitive specimen of the race, but on inspection, I cannot myself discover "within" me, a large and complex hierarchy of desires, each level attempting to influence the workings of the one immediately "below" and little plots and sub-plots going on, here, there and everywhere, to overthrow the dictates of the guy above (of course, since it possibly extends infinitely upwards, I may only be able to gain access to an infinitesimally small portion of it) - this kind of picture postulated by Frankfurt seems to have dubious resemblances to some traditional models of motivational processes as are criticised and parodied (admittedly somewhat harshly at times), by the likes of Ryle. The fundamental reason, however, why I am unhappy to talk of "desires" of any order higher than the second, is really the nub of the whole thing, i.e. the increasing levels of theoretical sophistication, as one moves "outwards" (I prefer this to "higher") from the arena of first-order desires, the greater levels of objectivity. I will attempt to make this clearer.

Before talking of this specific issue, it is in order, I think, to say a bit about objectivity, as characterised by Nagel. We each of us view the world from "within", from the inside, we all occupy a particular position in space and time, we all have highly specific desires, wants, whims and hankerings which, from the inside, appear to stand in no need of any justification. From this (Subjective) standpoint, all that our desires seem to need is satisfaction; from this standpoint, our desires, projects and purposes seem to be of immediate and utmost importance. However, unlike the rest of the animal kingdom, it seems, humans have the facility to transcend the subjective standpoint and see it from the "outside", i.e. they can see this subjectivity from a vantage point which encompasses and explains it (Objective). From this standpoint, we are, each of us, revealed to be no more significant than the next person, we are inconsequential atoms of history, our own desires, projects and purposes are of no great importance, and there is no particular

reason why my desires should be satisfied in preference to those of anyone else. And I don't think I can over-stress that it is only our ability to adopt this objective standpoint towards ourselves which allows there to be a free will problem at all: from the inside, it appears that I am the source of a great many of my actions but, once viewed from the outside, I become aware of myself as contained in a natural world of cause and effect, and it is then by no means so easy to see myself as the author of my actions. I also want to stress that there are not simply two radically different standpoints, the subjective, and the objective. Rather, it is a question of degree, and may usefully be viewed as a series of concentric circles: the first circle is the "least theoretical", the world of pure appearances (such as I depicted above), which can only be properly understood from the inside, the second is "more" theoretical, encompassing, explaining, and criticising this world of appearance, the third (in turn), encompasses, explains, and criticises the second, etc.

However, I cannot stress strongly enough that (as I've already talked of with regard to "explaining away" moves in empirical determinism/incompatibilism) there <u>is</u> a limit to how far humans are constitutionally capable of carrying on this process, humans can only reach out to a finite number of the concentric circles. (The idealist will want to claim that there <u>cannot be</u> a circle beyond the farthest one to which humans can constitutionally reach out,i.e. what there is must be capable of being apprehended by humans, whilst the realist will allow that there could be a circle beyond the farthest away one to which we can reach, which encompasses, explains and criticises the last one we can get to - indeed, for the realist, there could be any number beyond that again.) Humans are constitutionally limited in this respect: they can only go so far out and the game is up for them. We eventually arrive at a circle for which we can find no external justification, which must (as far as we are concerned) stand on its own feet. As we move outwards, we will arrive eventually at various critical procedures themselves, for which there <u>can</u> be no explanation, which must simply reflect our constitution, habits, practices and interests. Let me now attempt

to apply all of this to Frankfurt's model.

It seem to me that the correct picture is actually this: we find ourselves, at the most subjective standpoint, with a set of first-order desires, which appear to us simply to be in need of satisfaction e.g. we may want a drink, or a meal, or we may feel like listening to some music, or whatever. An agent may or may not reach the first circle outwards, the first standpoint of greater objectivity, which attempts to encompass, explain and criticise what appears to us from our most subjective vantage point. The wanton drug addict fails to reach even this position. For someone who does reach it, the desires which, from the most subjective standpoint, appeared only to be in need of satisfaction, then appear as something with an explanation from outside themselves, and an evaluation which will often depend very much on what the agent, rightly or wrongly, takes this explanation to be. An agent may or may not get to the next position outward, i.e. a position which encompasses, explains and evaluates the standpoint of second-order volitions: as I've already indicated, the agent in our "alcohol" example completely fails to get there, since it simply doesn't occur to him that what appears to him at the second-order level, the first circle of objectivity, could itself undergo any scrutiny or evaluation. It is given (for instance, that carnal indulgence for its own sake is a bad thing). There obviously is another level, however, when we get here, what has appeared at the second-order level appears in need of evaluation and explanation: for instance, how can I explain and evaluate my belief (say) that carnal indulgence for its own sake is a bad thing? Should I get to this position, I may discover that I did, in fact, have no decent reasons at all for holding various principles and beliefs which had previously appeared at the first level of objectivity, and guided my second-order volitions. My second-order volitions may therefore undergo change. However, I have already said that there is a limit to how long this process can go on, how far "outward" humans can go. I have already said that I am reluctant to talk of "desires" beyond the second-order (for reasons which I hope are now looking defensible) and indeed, it is becoming a bit difficult to envisage

what would be involved, and what it would be like psychologically, if we attempt to go to the level of objectivity beyond that which criticises and evaluates second-order volitions. Some sort of examination of these procedures of criticism and evaluation themselves? Maybe. But then, it seems, it is all over. Nowhere to run, nowhere to go. We cannot look for any more, in terms of explanations and evaluation, at this point. This is the fundamental reason why I can make little sense of Frankfurt's suggestion that we can go forever, if we like, forming desires of progressively higher orders. And, if we are to take our line on freedom from Frankfurt, it would seem that the person who enjoys the greatest freedom of the will is the person who can bring his actions into line with what he wants to do, from the vantage point of the greatest possible level of objectivity which humans can reach. And this does not have the unmanageable consequences it may appear to have: we do not have to take this giant leap outwards, every time we act, there will be arenas of action, regarding which we needn't step outside the most subjective of standpoints (e.g. "which kind of chocolate do I want?"). However, the extent and delineation of even this arena itself will have been determined from the vantage point of the greatest possible objectivity which we can reach. This is also why I am not happy with the idea put forward by Frankfurt of commitments 'resounding' through arrays of higher orders, when a desire gets decisive endorsement at a certain level. Frankfurt is, in a sense, I think, putting the cart before the horse here.

However, we still cannot avoid the sorry climax to all of this. Despite all my huffing and puffing to fill in the blanks in Frankfurt's effort, I think that it is really ultimately to little avail. The person who reaches the n^{th} circle of objectivity clearly has a different kind of existence to the person who never leaves the most narrow, subjective standpoint. However, if determinism is true, then it is causally determined that our first subject reaches the n^{th} level of objectivity and the second one gets nowhere, and so to say, (as Frankfurt would) that our first subject (assuming that he is also unimpeded) lacks nothing in the way of freedom, is to pointlessly shift the goalposts, to evasively define free will so

that something which is actually fundamental, i.e. the question of determinism, then looks only marginal. If determinism is true, then it seems to me every bit as unwarranted to ascribe free will to our first, ultra-objective subject than to our second ultra-subjective one; in such a world, after all, it is not up to anyone how objective their judgements are, or how much their actions are in line with those judgements. I would then regard it no more reasonable to hold the first subject responsible for his thoughts or actions than the second one - how either of them actually gets into the picture, in a way which makes in reasonable to attribute responsibility to them, is not clear to me. I may be caused to have even a very rich, fulfilling existence: however, if I am caused to have it by forces outwith my control, then ascriptions of free will to me strike me as nonsensical, as would the attendant responsibility ascriptions. I repeat the point I made earlier against Mackie, that the genesis of even the most 'objective', fulfilled, empirical, contingent self cannot be marginalised with regard to questions of freedom and responsibility.

Frankfurt indeed, admits the severing of the link between responsibility and his "free will" - this is surely altogether very odd since Frankfurt seems then to be saying that we may not even hold responsible for his actions the person who lacks nothing in the way of freedom. For all his conceptual juggling, it seems to me that problems of determinism simply can't be stipulated or defined away. Even at this level, the determinist threat is there, in as full-blooded a fashion as ever before. We have not really moved, I think, from square one. There still seems little reason to doubt that, if determinism is true, then there is no free will, and persons are not responsible for their actions. Perhaps Frankfurt simply wanted to draw attention to things he thought essential to the best of what life has to offer, and indeed, perhaps if someone has "free will" in Frankfurt's sense (or, at any rate, in the sense in which I attempted to rescue it), he may not feel especially concerned about whether he has free will in the more problematic, fundamental sense, whether determinism is true, whether he should actually congratulate himself about having these life-enhancing

conditions. Would anyone actually care, if in Frankfurt's sense, they "lacked nothing in the way of freedom"? These, however, are really separate issues; whether or not the absence of determinism is essential to the Good Life is one thing, whether or not determinism is true, and whether we are responsible for our actions, is another thing again. As I've said, despite all the huffing and puffing, this latter question is still there in all its grim reality.

As with conditionalism, there is nothing said by Frankfurt which counter-evidences Van Inwagen's First Formal (incompatibilist) Argument, nor anything which evidences free will, in the substantive sense which matters. If Van Inwagen's (M) hypothesis is true, then it is up to Martians that, for instance, Smith doesn't get "out" beyond the "wanton" level, whilst Jones becomes profoundly self-aware, so even Jones's profound self-awareness is in a very serious sense, not up to him. If (M) is true (for instance), then Smith has not got free will and Jones has not got free will; Smith is not responsible for his actions, and Jones is not responsible for his actions. Straight and simple. Indeed, whilst I just cast doubt on whether the falsity of (M) and all significantly similar hypotheses is actually essential to the Good Life, it would be very difficult, I think, to imagine it having absolutely no impact whatever on the outlook of an optimally "objective", fulfilled person, if they discovered that there were good grounds for entertaining (say), hypothesis (M). Not only, of course, would they withdraw responsibility-ascriptions from themselves, but (depending on temperament), they may or may not be affected by a measure of unrest, and maybe even panic.

I don't think that this reflection is gratuitous. On the contrary, I think it deeply revealing. Indeed, harping back to comments made right at the outset, I think this point also reveals a great deal about the primitive, preverbal worries at the bottom of the ado which is the problem of free will. It reveals, I think, that whilst we would find it futile to attempt saying precisely what we mean by "the self", it is at least a condition of the concept which we do have of it, that it does not sit comfortably alongside

the prospect of the truth of (M), Ginet's 'H' hypothesis, or anything significantly like those. That is, at the most primitive level there appear to be a network of inextricably woven, problematic and seemingly elusive notions, including those of the self, action and responsibility, and this network would seem to entirely collapse if (M), H or anything like it were true. There is clearly a lot we do not know about those notions, but this much we do seem to know. A recounting of those primitive intuitions makes it clear, I think, just how little help or comfort can be provided by what we have just seen Frankfurt do. Frankfurt seems to be attempting to persuade us that certain issues don't really matter, or ought not really to matter, but what is supposed not to matter is what is, unavoidably, of the most profound importance. If (M), or its like, were true, then whilst we may regard a preponderance of satisfied pigs as a regrettable state of affairs, it would be unwarranted to harbour feelings of distaste towards the satisfied pigs themselves, whilst a dissatisfied Socrates, whilst he may lament the state of affairs which leaves him dissatisfied, would be unwarranted in scowling at himself for failing to be perfectly rational, virtuous and happy. Even the smirk of a satisfied Socrates would become somewhat nervous, I would imagine, if he thought there were reasons for believing (M). Without wishing to stray too far here, whilst a unified theory of the Good Life seems hopelessly difficult, many would agree, I expect, that a component of the Good Life is the belief that one is, in some measure, to be genuinely praised and esteemed for certain kinds of achievement, acquisition, etc. And this would seem impossible if (M) or the like were true, quite regardless of anything offered by Frankfurt.

I talked a moment ago of concepts and conditions of action, the self and responsibility. It is by focusing on the first of these, the conditions of an act, that the next (final) and by far most difficult and challenging argument in favour of compatibilism which I will consider, gets off the ground. Later, I will consider a challenge to assumptions about conditions of responsibility, a challenge issued, in fact, by Frankfurt himself. The compatibilist argument, which I will now document and examine will plunge us into

several complex and difficult areas; some discussion of theory of action, for instance, will be essential and the full significance of issues touched on earlier, such as Van Inwagen's distinction between mere causation and determinism, and the Libertarian's Dilemma, will become clear, I hope. Van Inwagen has called the argument the Mind argument, owing to the number of times it has occured in the journal of that name, and I can see no reason to depart from this name.

10. THE MIND ARGUMENT

It is a fairly commonplace realisation amongst contributors that, whilst the truth of the thesis of determinism would at least problematise any belief that we have in free will, the falsity of the determinist thesis appears (prima facie, at least) to have similarly unnerving consequences. If determinism is true, then unless some form of compabilitism is correct, there seems no reason to hold me responsible for my actions, to regard them as mine, in any substantive sense. However, what if determinism is false? Can there, in such a case, be any better reason for holding me responsible for my actions, for regarding them as mine? It is not at all obvious, straight off, that there is. I cited an articulation of this worry near to the beginning, the articulation provided by Dennett in "On Giving Libertarians What They Say They Want". Dennett notes Hume's point that the dilemma posed for the libertarian is that, whilst determinism doesn't seem to sit too neatly alongside free will, if, on the other hand my actions are not determined by anything, then they appear to just happen at random, in which case there seems no reason either to hold me responsible for them, to regard them as mine. And it may be remembered that the task Dennett set himself was that if breaking down this dichotomy on the libertarian's behalf, of demonstrating that "undetermined" needn't mean "random", and consequently, that it may be possible to discover some area of indeterminism which would, in fact, help give credence to our notions of personal authorship and responsibility. Whilst (as it may also be recalled) I said that I was, largely, unimpressed by Dennett's specific efforts to this end, the worry which motivated his enterprise in the first place, is, at bottom, the same concern which generates the Mind argument in favour of compatibilism. Recognising (as Dennett has done) the problem which appears to be set for free will by indeterminism, the proponents of the Mind argument actually assert that determinism is a condition of an act. If we are to act at all (it goes), then determinism must be true.

The proper examination of this compatibilist position will involve us in a hornet's nest of issues. For instance, I have already touched on Van Inwagen's purported distinction between mere causation and full-blown determinism, and this will become of some importance, since it may well be that this compatibilist position is actually motivated through fears about the absence of causation with regard to actions, whilst this in turn can only be a worry about the absence of determinism if causation entails determinism. What if we could provide a solution to the 'causation' worry about actions which wouldn't entail the truth of determinsm? Whether we would then be any better off with regard to free will would depend on what form the suggested solution takes. In the same way that Dennett attempts (albeit fruitlessly, I believe) to deconstruct the supposed synonymy of "uncaused" and "random" to provide a path for freedom and reponsibility, Van Inwagen attempts to break down the supposed synonymy of "undetermined" and "uncaused" in the hope of demonstrating one of the confusions inherent in the Mind argument. Another difficult notion which it will be necessary to consider in connection with the Mind argument is that of agent-causation, an idea which allegedly dates back to Aristotle[1], is given a well-known modern defence by Roderick Chisholm[2], and gains little praise from Van Inwagen. Basically (for the time being), if agent-causation is true, then we are presented straight off with causation without determinism in relation to action, and therefore the refutation of the Mind argument.

Before getting our teeth straight into the argument proper, it is worth noting one other important feature of what I have said so far, i.e. that (so far, at any rate) the Mind argument appears to be only an argument that an act requires determinism - it says nothing about free action. Action may well require determinism, but extra arguments are then required to demonstrate that an act, so conceived, is then a free act. That is, as far as this goes, the proponent of the Mind argument may be entirely correct: if so, then determinism is a condition of action. However, this would be consistent with the absence of free actions from the world. All actions could yet be determined and unfree. More is required,

therefore, to say how an act, given that it is, qua act, determined, can be free. How would we distinguish, within the class of acts, between those which are free and those which aren't free? It cannot be by the presence or absence of determinism since all acts are, by hypothesis, determined. There would have to be some other criterion.

So the Mind argument, as I have portrayed it so far is not, strictly, a compatibilist argument. It is only an argument that action requires determinism, and more is needed to show that determinism is compatible with free action, and hence turn it into a compatibilist argument. For all we know at present, free will may be both incompatible with determinism (as I have myself argued so far) and incompatible with indeterminism, and hence impossible. I will come to this extremely difficult problem.

If my incompatibilist claims so far are sound, then a free act (if it is to be possible at all) must be an undetermined act. However, it is when we begin to properly consider what an undetermined act would really be like that we can begin to glimpse what is the force behind the Mind contention (leaving aside the question of the ultimate defensibility of it). This force is, remember, in a sense an essentially negative one, taking root not so much in a positive belief about the compatibility of free will and determinism, but in a converse worry about the alleged compatibility of free will and indeterminism, and indeed, according to one strand (as I've said) in a worry about the alleged compatibility of action and indeterminism. As I discussed in general terms away at the very beginning, and have given specific content to on occasions up to now, the legacy of Wittgenstein and the analytic tradition has brought about a temptation to regard the apparently most intractable of philosophical problems as cases, really, of much ado about nothing, turning upon simple failures (for instance) to be receptive to the nuances and "tricks" of our language. I hope that it is clear by now how problems of free will and determinism have been vulnerable objects of this approach, and I also hope, similarly, that some of the stage has been set for my saying that, in his articulation and setting out of problems within free

will/determinism, Van Inwagen displays a quite ingenious facility for exposing the limitations of this dismissive approach, and for holding up the traditional problems of the domain to be the perfectly real, substantive problems that they are. I say this now because I feel that his presentation at least, of the specific issue I am now about to discuss is again ingenious in its force and clarity. I often think myself that the most efficacious way there is of defending the authenticity of classical philosophical problems is actually to present logically posssible, even if especially fanciful and/or detailed and picturesque hypotheses, which appear, prima facie, to display the problem in all its reality, and then throw out the challenge as to why this would not be an articulation of a genuine philosphical problem. For instance, despite everything, despite, for instance, the efforts of Mind contributors like O.K. Bouwsma[3], I think that the most famous of these hypotheses, namely Descartes' Evil Demon, incontrovertibly reveals the problem of Scepticism to be an absolutely real one, and no amount of Moore-like manoeuverings[4] will make the problem dissolve into an empty one. The answer, whatever it is, is of course, a different matter altogther, but the point is that it is an answer to an actual, full-blooded, genuine problem. In our time (though this is not the place to talk about it), Thomas Nagel, in The View from Nowhere, has come up with a brilliant fantasy-hypothesis, which holds up the reality of the Realism/Idealism problem against its discreditors[5]. And of course, getting back to the issue I am specifically concerned with, Van Inwagen's Essay on Free Will, is permeated by his being in the same kind of role as, performing the same kind of function as, the likes of Descartes and Nagel in the other cases I have mentioned. His aforementioned (M) hypothesis is itself a superb example of this: it is a logically possible, though extremely fanciful, hypothesis which, if true, leaves determinism true and free will false. Straight and simple. The (M) hypothesis (and we can hardly forget Van Inwagen's statement of determinism, nor his First and Second incompatibilist arguments) is an exceptionally useful device with regard to demonstrating that, despite the smugness, the manoeuverings, or the writhings of the likes of Strawson, Sellars, Austin, Davidson, Dennett, Mackie,

Lehrer or Ayers, the problem of free will and determinism is a real and difficult one which won't go away simply by our being thorough Wittgensteinians (or convinced Kantians, for that matter). Or, at any rate, it demonstrates that one supposed problem within free will/determinism, namely the problem of compatibilism, is a perfectly real problem (and, of course, in the view of Van Inwagen and myself, that free will and determinism are incompatibile). Similarly, as I've said, with this specific problem within free will/determinism, the compatibility of indeterminism and free will, or as the case may be, simply the compatibility of indeterminism and action: any temptation that may exist to dismiss the supposed worry as a non-problem, as some failure to grasp the workings of our language properly or some equally primitive error will evaporate, I think, on looking at the splendid, concrete hypothesis with which Van Inwagen presents us. Like the cases of Scepticism, Realism/Idealism, and the compatibility of free will and determinism, the answer is, again, something else altogether, something quite over and above the clear statement of the problem. Indeed, I will ultimately have to voice some misgivings about Van Inwagen's approach to solving the problem of the compatibility of free will and indeterminism. However, what matters, for the moment, is his superb explication of the problems in question, and I will lean on these as I go along. Like the "evil demon", the (M) hypothesis, etc., they are supremely felicitous, expressive articulations of the preverbal, 'gut' worry which is that of the compatibility of free will and indeterminism.

Van Inwagen separates the Mind argument into three strands, which have a common beginning in a certain set of reflections on what the nature of free action must be if the incompatibilist is right. I pre-empted aspects of the Mind argument (as I noted a few pages back) much earlier when I spoke of Dennett as, basically a 'language strata' theorist, and documented and criticised his approach to the Libertarian's Dilemma, which really seems to be, at bottom, Van Inwagen's "first strand" of the Mind argument. However, let us look, for the moment, at Van Inwagen's excellent little fantasy-illustration of this strand. Imagine a hardened thief who

is in the act of lifting the lid off the poor-box in a little country church. He sneers and curses when he sees what a small sum it contains. Nevertheless, he reaches for the money. Suddenly, there flashes into his mind a picture of the face of his dying mother and he remembers the promise he made to her by her death-bed always to be honest and upright. This isn't the first time that this has happened in the same kind of situation, but he has always disregarded it. But this time he does not disregard it. Instead, he thinks the matter over carefully and decides not to take the money. Acting on this decision, he leaves the church empty-handed. We may suppose that this decision was undetermined. That is, we may assume that there are possible worlds in which things were absolutely identical in every respect with the way they were in the actual world up to the moment at which our repentant thief made his decision - worlds in which, moreover, the laws of nature (crucially) are just what they are in the actual world - and in which he takes the money. According to the proponent of this strand of the Mind argument, this kind of example suggests that if we look carefully at the idea of an undetermined act, we shall see that such an act could not be a free act. This is because what this example (according to the Mind theorist) seems to illustrate is this: if the incompatibilist's account of free action is correct, then a free act is an act that is undertermined by prior states of affairs. But an act that is undetermined is a mere random or chance occurrence, and a random or chance occurrence is hardly the kind of thing that could be called a free act.

As I've already suggested, this charge of the collapse of our actions into randomness in the face of the absence of determinism, is the same one which Dennett makes an attempt at rebutting in "On Giving Libertarians What They Say They Want". And, in a significant sense, Van Inwagen's general thrust is the same as Dennett's. That is, like Dennett, Van Inwagen is unimpressed by the claim that, if an act is undetermined, then it must collapse into randomness. (It's possibly worth pointing out that their general baggage is slightly different: Dennett seems to run together determinism and mere causation, opposing himself at times to the suggestion that an

"uncaused" act must be a random one, whilst, as I've said, crucial aspects of Van Inwagen's position will, as we will see, turn on the alleged non-synonymy of "caused" and "determined". However, I do think that what is important about their respective positions on this point is the same, namely that it is not obvious that we are faced with having to make a straight choice between determinism and complete collapse into randomness.) As I devoted some time to defending, I am largely unimpressed by Dennett's specific efforts at utilising the possibility of a third option on behalf of free will; Van Inwagen's efforts towards the same end are, as I've said, of a different kind completely, focusing more on the fine-grained, conceptual distinctions such as that between determinism and mere causation. I will go on to attempt a serious appraisal of Van Inwagen's specific efforts at finding and utilising a third option on behalf of free will; perhaps they will carry more weight than I think Dennett's do. In the meantime, Van Inwagen has what I think are shrewed comments to make on the general premise that if our acts are undetermined, then they are mere "random" or "chance" events. The main problems are really of a conceptual nature, centering on what is actually meant by "random" or "chance". As Van Inwagen says, the words 'random' and 'chance' most naturally apply to patterns or sequences of events. So it is not clear what they might mean when applied to single events. These words might simply mean 'undetermined'; but, of course, in that case we should have no argument but only an assertion that undetermined events are not the sort of thing that can be called free acts. There is also a point which can be usefully documented from developments in computers, in order to rebut another familiar suggestion in this area. This is the suggestion that if our acts were undetermined, they would issue from us in a meaningless and incoherent jumble, as if we were perpetually deciding what to do by consulting a table of random numbers. However, we now know that (for instance) there are computers that sometimes change state in ways that are not determined by their earlier states and their input but their output is not random in the way in which a table of random numbers is random. The "first strand" is therefore looking decidedly dodgy; on these grounds alone, there is no reason to believe that it is a

straight fight between the determined, and some disordered chaotic morass. As I've said, Van Inwagen's attempt at exploiting the prospect of some or other third alternative will, to a large extent, depend upon his claim of the non-synonymy of 'caused' and 'determined', and will involve us in a difficult and complex set of issues. In order to get to them, we will have to move on to what Van Inwagen calls the "second strand" of the Mind argument, somewhere where the invaluable graphic and illustrative talents of Van Inwagen's which I have spoken of, will, I hope, be seen to excellent effect.

Before moving right on to this however, it is, I think, worth pausing on something for a moment. I have already said that, although it tends to be thought of as a compatibilist argument, the Mind argument is, strictly, an argument for the incompatibility of free will and indeterminism. Van Inwagen is as aware of this feature as I am, but nevertheless, he seems to omit the consideration of what I think are highly significant questions with regard to how the Mind argument fares as a contribution, specifically, to the compatibilist cause, and questions which he does indeed pose regarding other compatibilist arguments. For instance, does the Mind argument falsify Van Inwagen's First Formal (incompatibilist) argument? I would also myself reinforce this general point by asking (of <u>any</u> argument supposed to be supportive of compatibilism): if Ginet's 'H' hypothesis were true, would the argument (e.g. the Mind argument) help us make any sense of the notion of free will, of the possibility of our acting otherwise? It could well be that some or other strand of the Mind argument is a good case against the compatibility of free will and indeterminism (and I will attempt to arrive at a conclusion on this), but unless it can, for instance, come in the way of Van Inwagen's First Formal Argument, then there seems no reason to regard it as a good positive argument for compatibilism, i.e. for the compatibility of free will and determinism. Applying this criteria (as has been done with previous alleged compatibilist arguments), I do not see that the "first strand" of the Mind argument does anything whatever to discredit Van Inwagen's First Formal Argument, or to give sense to

the notion of free will in the face of the truth of Ginet's 'H' hypothesis. So, quite regardless of its consequences vis a vis the compatibility of free will and indeterminism, I don't see that the "first strand" could possibly be a positive argument in favour of compatibilism, a genuinely compatibilist argument. Even if it did go some of the way to showing that free will is impossible, it could hardly follow from this that it showed the truth of compatibilism! It is essenial, I think, that the same test is also applied to the "second" and "third" strands, otherwise there seems little way of knowing whether they are sound, positive arguments for compatibilism, or whether, on the contrary, their success would simply entail the impossibility of free will. Finally, in conclusion to the preamble, I feel it essential to point out that, whilst their precise import is different, the discussions of the second and third strands of the Mind argument are by no means entirely independent of one another; indeed, one specific issue (which I have already alluded to on several occasions), namely that of whether causes determine their effects, will be central to the discussion of both strands. This will, I hope, become entirely clear as I proceed.

The second strand of the Mind argument is basically this: if an act, or what looks superficially like an act, is not determined to occur by prior states of affairs, then it is not really an act at all. Whatever else an act may be, it is a production of its agent. But if an "act" (the argument goes on) is undetermined, it is not a production of its putative agent and hence not really his act at all. We can go back and work through what I've already said to be philosophically invaluable working examples, and what I've said Van Inwagen is so ingenious at providing. Let us go back to Van Inwagen's thief who refrained - or so we should say if we went by outward appearances - from robbing the poor-box. His refraining, or the event that we should initially be inclined to call his refraining, was ex hypothesi undetermined. Van Inwagen decides to assume psycho-physical identity as an aid to his presentation, but I think that mere psycho-physical correlation a la Honderich, would suffice. We can, I think (as Honderich does in "One Determinism")

remain neutral as to whether this psycho-physical correlation is in fact, a relation of identity, or whether it is some other (weaker, I suppose) relation. I don't see myself that the point would be made any less forcefully and, in highly contentious areas such as this, it is always very advisable, I think, to assume only what is absolutely essential as regards the point at issue. The assumption of psycho-physical correlation is, I think, a significantly lesser liberty than an assumption of psycho-physical identity, and (as I've said), I don't see anything being lost here through this particular exercise in intellectual humility. So, let us suppose psycho-physical correlation (or The Correlation Thesis).

Suppose that some event in your brain is correlated with one of your acts. And suppose that this event was undetermined by earlier events in your brain. Van Inwagen calls this undetermined event "E". There would seem to be no reason to suppose that E or any other undetermined event is <u>essentially</u> undetermined, and indeed this reflection is the first in a chain of reflections which are crucial to this aspect of the Mind argument, whose defensibility will turn out to depend upon the credibility (or otherwise) of this set of reflections. Van Inwagen provides this example: suppose that a cup sitting untouched on a table suddenly breaks at t and that this event - the cup's breaking at t - is undetermined by earlier states of affairs. It would seem that it <u>might</u> have been determined by earlier states of affairs: a hammer might have struck the cup at just the proper moment that its breaking at t should have been causally determined by this blow. (We can, I think, waive the lurking problem concerning event-identity. It will do to note that if an undetermined event happens, it is logically possible for a <u>descriptively</u> identical determined event to have happened at the same moment.) So we can suppose that an event just like E and having the same consequences had happened, and that it was determined. Now I will borrow from what is probably Van Inwagen's most imaginative and useful illustration contained in his <u>Essay on Free Will</u>, his Cartesian-style freakish demon. That is, we can suppose that the action in question was determined by a freakish demon. So, when all is said and done, we have the following

sequence of events: the freakish demon performs some supernatural act such that it is causally impossible for this act to be performed and E - or an event just like E - not to occur. Now let us suppose pulling everything together, that E is correlated with (in the example) an event we also call 'the thief's deliberations', under which description we include the outcome of these deliberations, this outcome then determines that the thief shall refrain from robbing the poor-box and shall depart. In such a case, it would seem that the thief would not really have acted at all, because he would not have been the producer of the event in his brain that initiated the bodily motions characteristic of a man leaving a poor-box empty handed. As Van Inwagen says, this case would not be significantly different from the case of someone whose brain has been "wired" by a brain-physiologist and whose arm rises whenever the physiologist presses a certain button. Indeed, although Van Inwagen does not himself remark on this, it is not significantly different from what would be the case if the (M) hypothesis were true; in the latter case, it would be Martians who were the producers of the event in the thief's brain which was efficacious, and in the case now under consideration, the Martians are simply replaced by a freakish demon. Now we may restore the "actual" situation: we remove the demon and his works, leaving, it seems, an uncaused event and its consequences. The crucial question is this: does this somehow change the "demon" story to a story in which the thief does act? And of course, it is very difficult to see how it could. The significant fact about the "demon" story is that it brings into prominence the fact that the determinants of the change are not to be found within oneself. And if an event is undetermined, then it seems just as true that the determinants of that event are not to be found within oneself as it would be if these determinants were to be found in the acts of a demon , for if the determinants of one's act are not to be found, then, a fortiori, they are not to be found within one. (Perhaps some small inkling is now being felt of how it could make a substantial difference whether one runs together causation and determinism, like Davidson and Hume, or whether, like Van Inwagen or Searle, one thinks it unwarranted to do this. At any rate, I hope the full significance of this will

become clear fairly soon.) Indeed, in order to reinforce this point, and to move some way towards discovering what form a possible solution might take, Van Inwagen embellishes the freakish demon story to (I think) splendid effect. There are some highly significant moments in the embellishment, significant in that they raise issues which it is essential that I confront.

Let us suppose that the demon exercise his influence over the thief in this way: there is an invisible wire that passes through the thief's skull and into his brain; at the other end of the wire is a sort piano that is made of subtle matter and upon the keyboard of which the demon plays; by what he plays he can "direct" via the wire the motions of the atoms in the thief's brain and can thereby direct the thief's inner life, including his deliberations. And this is just what he has done in the imagined case: by striking the keys in a certain order, he has guided the thief's brain through just that sequence of states that correspond to the deliberations of a man who refrains from stealing. Let us now, gradually modify things, and move towards what might be key factors. First, spice it with the supposition that the demon's actions at the keyboard are undetermined by the demon's own inner states and by anything else. It is important to be clear that this change in the case does not weaken the argument for the earlier conclusion (that the thief did not act). Now let us remove the demon and suppose that the "sort of piano" we imagined the demon to be playing is a sort of player piano: let us imagine that the keyboard is worked by a mechanism internal to the piano; and let us suppose that this internal mechanism is an indeterministic one. Nor does this change in the case weaken the argument for the conclusion that the thief does not act. Next stage: remove the piano and suppose that impulses simply appear in the wire (which now protrudes from the thief's skull, its far end unattached to anything) undetermined by prior states of affairs. Nor, again, does this change weaken the argument for the conclusion that the thief did not act. We are moving ever nearer to the punch-line, which the more adept may be able to guess. Let's imagine the wire becoming shorter and shorter till only the part inside the thief's skull remains. Again, how could this change

weaken the argument? If impulses undetermined by past states of affairs appear in the wire and if these impulses determine the putative acts of the thief - determine what movements are made by his limbs and what thoughts pass through his mind - then these putative acts are not real acts, are not among the things he does or produces. We can keep going: let us suppose that the wire is replaced by one of solid flesh, by a "wire", or wire-shaped thing, made of brain cells. Does this change succeed in weakening the argument? It would seem not: what difference does it make what the wire is made of? The important thing is that the impulses of which it is the carrier should be undetermined by past states of affairs. I will now deliver the punch-line. We can and might as well go the whole hog: let us suppose that the wire-shaped thing made of brain cells is a natural part of the thief's brain. It is not obvious, straight off, that even this modification should make a difference. After all, whilst it is indeterminism with regard to actions we are concerned with here, the determinism of the likes of Honderich or Ginet features (unlike the (M) hypothesis) entirely natural, non-freakish phenonema, yet this feature doesn't seem to make it any easier to see how their truth could be squared with free will. Let us look again, for instance, at Honderich's ('One') determinism:

> States of the brain are...effects, the effects of other physical states. Many states of the brain, secondly, are correlates...States of the brain, thirdly, are causes, both of other states of the brain and also of certain movements of one's body. The latter are actions...It follows from these three premises, about states of the brain as effects, as correlates and as causes, that on every occasion when we act, we can only act as in fact we do. It follows too that we are not responsible for our actions, and, what is most fundamental, that we do not possess selves of a certain character.

It seems clear from this, that where the issue of determinism and our implications in our own actions is concerned, it matters not one iota whether the determinants are Martians, as in the (M) hypothesis, or whether they are straightforward, natural brain states as in

Honderich's determinist hypothesis (and pretty much the same would be, at bottom, true, I think, in the case of the truth of Ginet's 'H' hypothesis). This is because what really matters (as has already been noted) is the relation between actions and past states of affairs. And so why should it be any different in the case we are now considering, i.e. the case in which my actions are undetermined by past states of affairs? As with determinism, it would seem that all that matters is this relation (or non-relation, as the case may be) between action and past states of affairs. The thief in Van Inwagen's example has a wire-shaped thing made of brain cells within his brain, in which there arise impulses, undetermined by past events, which determine his every action. But, as Van Inwagen notes, if we now suppose that this thing is a natural part of the thief, we are making a supposition not about its momentary operations, but only about its origins. But facts about the origin of a part of one's brain can be relevant to questions of free will, it seems, only in so far as they are relevant to questions about the momentary operations of the part.

So we are now at the point in Van Inwagen's illustration, of imagining the thief's "acts" to be the result of undetermined impulses originating in a certain section of his brain. The freakish demon and his apparatus has gradually been turned into a natural part of the thief's brain. And in the final story, as in the original story, it still looks as though the thief does not act. So it looks like the whole idea of a undetermined act is incoherent and a fortiori, a free act (if it is to be a possible act at all), cannot be an undetermined act. The possible means which Van Inwagen floats of escaping this difficulty, and which I will try to appraise, will involve us in the complex hornet's nest of problems which I have been anticipating up to now.

Before getting involved in this, however, I feel inclined to ask at this point, the required set of questions regarding this strand of the Mind argument in so far as it purports to be a positive argument for compatibilism. Does it falsify Van Inwagen's First Formal (incompatibilist) argument? If (say) Ginet's H

hypothesis were true, would the soundness of this second strand of the Mind argument help us give genuine content to the idea that we can act otherwise? If Van Inwagen's own (fantasy) (M) hypothesis were true, would the success of the second strand of the Mind argument preserve free will nevertheless? On all counts the answer seems to me to be a flat "No". It does not seem to me that the second strand of the Mind argument contributes one iota to the positive cause of compatibilism, to squaring free will with determinism. If it turns out to be sound, then it would seem to me to demonstrate absolutely no more than that acts, qua acts, are determined; it does nothing for the separate cause of showing how (determined) acts could be free, of how we could act other than the way we do, in fact, act. (As will be seen, parallel questions will arise with the case of the Mind argument's third strand: it purports to show that the idea of an undetermined free act is an incoherent one, but is anything done in the process to help make sense of the notion of a determined free act?) Van Inwagen himself doesn't seem to be especially keen to hold it up to the light, but I feel it of some importance to say now that it looks very much to me that, if the second strand of the Mind argument can be sustained, then free will just looks impossible. The game would be up. Acts would have to be determined, but since determinism is incompatible with free will, then free acts are impossible - we can never act other than how we do act. That is how it looks to me. Of course, perhaps this whole conclusion (the impossibility of free will) can be averted; perhaps there are flaws in the second strand of the Mind argument, just as there seemed to be in the first strand. It is now time to document and appraise possible ways of dealing with this problem, and it is useful for me to repeat at this point that the second and third strands of the Mind argument and ways to coming to terms with them, are not entirely independent (as will become clear, I hope). Indeed, one general set of considerations are central to any attempted solution to either. This should be no surprise, of course, since we are dealing with strands of what is, really, one general argument. Indeed, it should also on this ground, be no surprise, that the form the solution (or the attempted solution, at any rate) to the problem will take has already, really,

been outlined in connection with the first strand of the argument. What really is generally required for any sort of solution to the Mind argument (at the very least) is the deconstruction of an alleged dichotomy: the alleged dichotomy of determined and random, meaningless. As I said, this was seen in connection with the first strand (and as I also said, we saw Dennett's attempts, much earlier, at confronting what is, at the bottom, the same supposed dichotomy), and the second strand seems permeated with the same supposition on a much more macroscopic scale. The second and third strands are really a kind of "blowing-up" of the picture of the first strand, and this has the advantage of allowing us to see more clearly, in more fine detail, exactly what is going on and what may be wrong with it. We must, in parallel, attempt to "flesh out", so to speak, the general counter-argument that we do not, despite appearances, have to choose between full-blown determinism or a complete collapse into randomness. And the "blown-up" picture afforded in the shape of Van Inwagen's imaginative fantasy - sliding into possible reality, may well allow us to give some substance to the general form of our counter-argument.

Van Inwagen suggests that it may not, after all, be an entirely harmless step in the progression, to move from "wire-shaped thing" to "natural wire-shaped thing". to suppose that the wire-shaped thing is a natural part of the thief. It will be recalled that I noted a few pages back that the shift from (say) Martian manipulation to entirely natural (neurophysiological, really) factors didn't seem to matter in the case of determinism. However, the supposition that the equivalent shift within indeterminism is equally innocuous is not only disputed by Van Inwagen, but indeed, this particular moment is the point of impact at which (if you like) the spirit of the solution to the Mind argument can, in the second strand, become flesh. That is, what is not clear to Van Inwagen is the claim that an apparent act of a human being that was the consequence of an undetermined change in a natural part of a human being could not be a real act, and it is through closing in on this point that Van Inwagen is both allowed and required to give substance to his disclaiming of the "determined/random" dichotomy.

That is, as with all aspects of the Mind argument, qua Mind argument, it is an absolute requirement to any solution that the essential dichotomy is deconstructed, and it seems to Van Inwagen that it is through moving the pointer to the apparently innocuous step I have cited that it becomes clear how we could begin to do this. Let me recap: what is required is that it is somehow made possible to view human action as undetermined, yet not random. Indeed, we want to discover if it is at all possible to view my "actions" as genuinely mine. If my incompatibilist position put so far is correct, then this isn't possible if determinism is true. So we are looking for the possibility of "actions" being undetermined, yet still occurring in some or other sort of way which grounds my "actions" as really mine, as real acts (and ultimately, as real, free acts).

One way of approaching this may be to enquire as to the relation between determinism and causation; for instance, what if we just could make some sort of sense of the notion of an "act" which is undetermined, yet which is not random, but indeed is caused in such a way as to ground the belief that it is a real act? This is precisely the possibility which Van Inwagen wishes to explore. His general disputing of the dichotomy which runs through the Mind argument is given flesh in this case in his contention that we can make sense of the idea of causation without determination and that, where we are concerned with an undetermined change in a natural part of a human being, then we may just have a case, not only of causation without determination, but of a causation which grounds the notion of genuine action. (As I've already indicated, this idea will also be important within the third strand of the Mind argument, the worry about undetermined free acts.) I have, in fact, already alluded to Van Inwagen's (not especially widely accepted, it seems) distinction between mere causation and full-blown determination: away at the beginning, I cited Van Inwagen's distinction between the thesis of determinism and the Principle of Universal Causation. It will be remembered that determinism was defined by Van Inwagen as the thesis that there is at any instant exactly one physically possible future and that in order to deduce determinism from the

Principle of Universal Causation we should need at least three premises. It is worthwhile, I think, to re-state one of those premises at this point since (as I said then), the truth alone of one doctrine I will devote some time to soon, namely agent-causation, would falsify it straight off, giving us a case of causation with determination:

> if an event (or fact, change, state of affairs, or what have you) has a cause, then its cause is always itself an event (or what have you) and never a substance or continuant, such as a man.

'Agent-causation' is a possible means, one of two which are largely recognised by Van Inwagen, by which our hope may be realised i.e. undetermined causes which ground genuine action (and later, free action). Both possibilities which Van Inwagen recognises, as I've said, home in on the idea of undetermined changes in a natural part of a human being. I am in agreement with Van Inwagen that the form any general solution must take is that of finding a causation without determination which is such that it grounds real action (later, free action); I am also in agreement that the standard Humean view of causation (which would preclude causation without determination) seems misguided, that tenability can be given to the idea of causation without determination, though (as I will make clear), I am sceptical as to whether Van Inwagen's specific efforts to this end are actually fruitful. Despite the seemingly widespread appeal of the Humean view, considerations against it are put forward, not only by Van Inwagen, but also by John Searle, in his thorough book, Intentionality, and I will conclude that the comments of Searle are much more persuasive than those of Van Inwagen. Before going on, I say again that I will eventually have some unease about what finally emerges from Van Inwagen vis a vis the Mind argument, worries which I think may be good reason for taking a serious look at the notion of agent-causation.

Before saying much about the Humean view of causation, I will provide a relatively brief sketch of the notion of agent-causation.

As I've indicated, the truth of agent-causation would entail the falsity of the Humean view, and solve a great many problems in this area. I will then move on to examine the Humean view itself, and consider how its falsity might help ground (as well as agent-causation) yet another model of human action, the one which is, at bottom, the one which Van Inwagen leans towards. I will eventually move full circle, coming back round to see whether agent-causation, or something significantly similar to it, can bear any fruit. It may well be that, when all is said and done, we will have to be content in an area so complex and at times seemingly intractable as this, with what is merely the _least repugnant_ of all the apparent alternatives.

The medieval philosophers believed that God was the cause of all things other than himself. However, most of them, at least, also believed that God was changeless. This raised a problem, since the causes we observe in nature produce their effects only at the expense of some change in themselves. Thus the Schoolmen postulated a type of causation that does _not_ involve a change in the cause and called it _immanent_ causation. In "Freedom and Action", Roderick Chisholm revived this idea and applied it to the human agent, Richard Taylor later coining the term "agent-causation" for the same thing. According to Chisholm and Taylor, the _agent himself_ is sometimes (when he acts freely), the cause of his own acts or, perhaps, the cause of the bodily or mental changes that manifest them. This doesn't mean that a change in the agent causes these acts or changes; rather the agent _causes_ without himself changing in any way. I hope it is very easy to now see that if the thesis of agent-causation is true, then an act can be undetermined. Going back to the case of Van Inwagen's thief, to say that it was not determined that he should refrain from stealing is to say this: there is a possible world that (a) is _exactly_ like the actual world in every detail up to the moment at which the thief refrained from stealing, and (b) is governed by the same laws of nature as the actual world, and (c) is such that, in it the thief robbed the poor-box. (We will, of course, see exactly this kind of possibility appear again in the model which Van Inwagen leans towards, and an

attempted explanation of how it could be true.) If agent-causation is true, then it seems that these conditions (of indeterminism) are satisfied, i.e. that whilst the thief refrains from robbing the poor-box, there is nevertheless a possible world that is exactly like the actual world in every detail up to the moment at which the thief refrained from stealing, is governed by the same laws of nature as the actual world, and in which the thief robbed the poor-box. The causally relevant difference between the two given worlds is neither some or other set of events in the past nor some or other set of laws of nature (since both are identical across the two worlds), but what is caused by the agent at the crucial moment. Therefore, if agent-causation is a coherent concept, the second strand of the Mind argument fails to show that the notion of an undetermined act is incoherent. I will leave this notion, now that some idea of it has been given, for the time being, except to reiterate, not only that if it is true, we have, straight off, a refutation of the (standard) Humean view of causation, but also that the falsity of this Humean view may well facilitate (as Van Inwagen believes it does) some other non-deterministic view of human action and another view which, for all we know at present, may be more plausible than agent-causation. However, we must examine the Humean view on its own merits; it would be a bad idea to wait for an enquiry into agent-causation, and hold our breath on the Humean view. For all we know at present, the Humean view could be false, agent-causation false, and some other non-deterministic model of human action correct. (And as I've indicated, Van Inwagen's sympathies lie towards these respective truth-values.) I will now look at the Humean view of causation, the view that causation and determination are necessarily run together, entirely on its own merits.

We have, of course, already seen what is probably the most well-known (and it seems, generally accepted) feature of this Humean view of causality, in the shape of Davidson's assumption of the Nomological Character of Causality. It is slightly ironic that Van Inwagen for one, expresses sympathy with Davidson's general theory of action, yet spends some time disclaiming Davidson's Humean stance

on causation (which, I suggested, forces him to espouse a straightforward physical determinism, at bottom, and consequently close the door on free will). It will be remembered also, I expect, that when I discussed the 'language strata' theorists, of whom Davidson is the most distinguished, I mentioned in passing the contribution of John Searle to a family of questions which are now highly relevant and requiring detailed examination, concerning causation, intention and action. I have, in some sense, moved round the circle, and have come to the point where some ideas merely floated at that point have to be properly cashed out and discussed, and where, as I've said, it is of importance to examine Davidson's Humean assumptions on causality, as well as more general features of his theory of action (which I did, in fact, express some sympathy with earlier e.g. intentional states as causes of action).

Whilst Searle and Van Inwagen tend to approach things from opposite directions, their efforts tend to complement one another. Van Inwagen seems to make it a more primary matter to criticise the general Humean ideas on causation, thus facilitating a non-deterministic version of Davidson's theory of action (which he leans towards), whilst Searle focuses more primarily on action itself, and its relation with intention, a proper understanding of which, he thinks, discredits straight off some Humean ideas on causation, as well as grounding a theory of action which seems very much like that favoured by Van Inwagen (and Davidson, of course). I will, as I said, be expressing doubt about Van Inwagen's efforts at discrediting the Humean picture (though I am, in fact, sympathetic to the anit-Humean conclusion). Searle's efforts to this end are, I think, much more compelling. And as I proceed, I will lean fairly heavily on the contributions of both of them, part of which (at least) are, I think, highly shrewd and penetrating.

Whilst Van Inwagen concerns himself, really, with only one feature of this accepted view of causation, Searle documents and sets about criticising three main features of it. The criticism of all three features will be seen eventually, I hope, to be of relevance, and indeed (as I've said) to be of some help for Van

Inwagen's purposes. Searle cites the overworked philosophical example, the paradigm case of causality: billiard ball A makes its inevitable way across the green table, where it strikes billard ball B, at which point B starts to move and A ceases to move. The event of A's striking B caused the event of B's moving. The three main features, given by Searle, which this little affair supposedly instantiates, are:

(1) The causal nexus is not itself observable. One can observe causal regularities; that is, one can observe certain sorts of regular sequences of events in which events of a certain type are followed by events of another type; but in addition to regularities, one cannot observe a relation of causation between events. In the way that I can literally see that the cat is on the mat (Searle's example), or that one event followed another event, I cannot literally see that one event caused another event. In the billiard ball example I see events which are in fact causally related, but I don't see any causal relation in addition to regularity.

(2) Whenever there is a pair of events related as cause and effect, that pair must instantiate some universal regularity (Nomological Character of Causality). For every individual case where an event causes another event there must be some description of the first event and some description of the second event such that there is a causal law correlating events fitting the first description with those fitting the second.

As Searle notes, however, it is important to distinguish between the metaphysical and linguistic versions of this "regularity" idea. In the metaphysical version every particular causal relation is <u>in fact</u> an instance of a universal regularity. In the linguistic version it is part of the <u>concept</u> of causation that every singular causal <u>statement</u> entails that there is a causal <u>law</u> correlating events of the two types under some description or other. The linguistic claim is stronger that the metaphysical claim in the sense that it entails the metaphysical claim but is not entailed by it. (Indeed, it will

be seen, as I go on, that Searle himself believes the linguistic thesis to be false, and the metaphysical thesis most likely to be true.) In the Davidsonian version it is also claimed that to know that A caused B, one must know that there is a law. For instance, Davidson writes

> In any case, in order to know that a singular causal statement is true it is not necessary to know the truth of a law; it is necessary only to know that some law covering the events at hand exists6

Anyone familiar with Davidson's offerings on this and related topics will know that Davidson is also one of a host of philosophers (the aforementioned David Pears is another) who suggest that the law need not be stated in the same terms as the singular statement. I cited this point in my specific discussion of Anamolous Monism, and its consequences for free will.

(3) Causal regularities are distinct from logical regularities. There are many regularities that are not even possible candidates for being causal regularities because the phenomena in question are logically related. Thus, for example (Searle's) being a triangle is always associated with being three-sided, but something's being a triangle would never cause it to be three-sided since the correlation is by logical necessity. The aspects under which one event causes another event must be logically independent aspects. Again, this metaphysical thesis has a linguistic correlate in the formal mode. The causal law must state regularities under logically independent descriptions, and therefore must state a contingent truth. ||It will be specifically remembered, I hope, that in my discussion of Davidson and the 'language strata' approach, I touched on the apparent problem thrown up for a (Davidsonian) causal theory of action by the seeming contingent nature of causal relationships, and the seeming a priori relationship between purposive states and actions. I floated a solution at that point, in barest outline, and said that the contribution

of Searle would help significantly to flesh it out. This will be on the agenda fairly soon.

As I said, through Searle's attacking this entire picture of causation, an alternative position to a deterministic view of human action is made possible. However, as I've been repeating all along, we could be well on the way to solving some problems if we could even just dismantle the second feature of the standard picture of causation, as cited by Searle. Before documenting reasons for being sceptical of it, however, it is worth taking note of how Van Inwagen cites the same premise:

> Many, perhaps most, philosophers believe that causes <u>determine</u> their effects. They believe, that is, that given the cause...a relation that takes <u>events</u> or <u>states of affairs</u> or some such, and not <u>persons</u> (i.e. not 'agent-'causation), as it terms, the effect <u>must</u> follow. They believe that if A is the cause of B, then, if A had happened and had <u>not</u> been followed by B, this could only have been because the laws of nature were different.[7]

As I've indicated, neither Van Inwagen nor Searle believe that there is very much to commend this premise, and I think that Searle at any rate adduces some compelling considerations against it. As I've also indicated, some of Searle's considerations, not only undermine the general premise, but actually flow from an independently supportable non-deterministic (though causal) account of action. Let me first cite the more general example of Van Inwagen's, an example designed to generally counter-evidence the (Humean) premise that causes determine their effects. As I've indicated, whilst I think there are good grounds for doubting this Humean picture, I'm ultimately sceptical as to whether this Van Inwagen example provides any such grounds.

Suppose someone throws a stone at a window, and that the stone strikes the glass and the glass shatters in just the way we should expect glass to shatter when struck by a cast stone. Suppose further that God reveals to us that the glass did not have to

shatter under these conditions, that there are possible worlds having exactly the same laws of nature as the actual world and having histories indentical with that of the actual world in every detail up to the instant at which the stone came into contact with the glass, but in which the stone rebounded from the intact glass. It follows from what we imagine God to have told us that determinism is false. But (Van Inwagen poses) does it also follow that the stone did not break the glass, or that the glass did not break <u>because</u> it was struck by the stone? Van Inwagen says that the only reason we could have for saying this is that we accept a corollary of the standard theory of causation: that instances of causation simply <u>are</u> instances of universal exceptionless laws, that the concept of the instantiation of an exceptionless law and the concept of causation are one and the same concept. With his hitherto well-documented flair for fantasy-examples, Van Inwagen extends this case. Suppose that we are watching a slow-motion film taken by a camera trained on the point at which the stone came into contact with the pane of glass. We observe the following: the stone moves through the air towards the point of impact; the stone touches the glass; the glass bends ever so slightly; cracks appear in the glass, radiating outward from the point of impact; shorter cracks appear, joining these radial cracks, thus producing detached shards of glass; the stone moves through the space formerly occupied by the unbroken pane, pushing,or apparently pushing, shards of glass out of its way; the stone continues along its path, trailing a rough cone of spinning shards. It seems trite to say (as Van Inwagen does) that watching this slow-motion film several times should make it very hard to believe that the stone did not break the glass. Van Inwagen then asks us to harp back, for the moment, to the revelation we earlier imagined God having delivered:

> Could this revelation really lead us to say that, despite appearances, the stone didn't cause the glass to break? That it is a <u>logical consequence</u> of this revelation? Wouldn't it be more reasonable to say this: that, while the stone did cause the window to break, it was not <u>determined</u> that it should; that it <u>in fact</u> caused the window to break, though, even if all

conditions had been precisely the same, it might not have?[8]

As I indicated, I do myself feel a problem with this attempt on Van Inwagen's part at refuting the Humean assumption, and I think that what worries me is that it looks question-begging. We are asked to imagine that we have compelling grounds for believing that the glass broke because it was struck by the stone; Van Inwagen's fantasy slow-motion film would seem to provide such grounds. Fair enough. However, we are also asked to imagine God revealing to us that the glass did not <u>have</u> to shatter under these conditions, that there are possible worlds having exactly the same laws of nature as the actual world and having histories identical with that of the actual world in every detail up to the instant at which the stone came into contact with the glass, but in which the stone rebounded from the intact glass. This is precisely what seems question-begging to me, i.e. it is <u>assumed</u> here precisely what Van Inwagen has to demonstrate, namely that there <u>is</u> a possible world having exactly the same laws of nature as the actual world and having histories identical with that of the actual world in every detail up to the instant at which the stone came into contact with the glass, but in which the stone rebounded from the intact glass. If there <u>is</u> such a world then, of course, determinism is false, the (Humean) assumption that causes determine their effects is false, and (what really matters here) the door is opened to undetermined causes and possibly to undetermined causes (e.g. in a natural part of a human being) which will ground genuine action, and maybe even genuine free action. But we can hardly just assume that there is such a world; for all we know at present, there might not be - it might be the case that the Humean position is correct, that causes <u>do</u> determine their effects, that in <u>all</u> possible worlds having exactly the same laws of nature as the actual world and having histories identical with that of the actual world in every detail up to the instant at which the stone came into contact with the glass, the glass shatters when struck by the stone. As I've said, it is incumbent upon Van Inwagen to demonstrate that this is not the case. Instead, it is assumed that it is not the case. Perhaps the idea of God, and his revealing things to us, confused Van Inwagen somewhat. This is not

the place for a discussion of the logical limits of God's omnipotence, but I feel bound to make a few points.

Whilst it would, I expect, be uncontroversial to say that God may reveal possible worlds to us which may otherwise remain hidden from view, there are surely a set of significant questions, which Van Inwagen overlooks, concerning God's power of creation of possible worlds, and how this stands in relation to his supposed revelations. Let me clarify. I have already noted the distinction Searle makes between the linguistic "regularity" thesis about causation and the metaphysical "regularity" thesis. Similarly, not only is it generally thought to be essential to distinguish between logical (or 'analytic') and factual (or 'synthetic') possibilities, but this distinction has also had considerable import within enquiries about God's omnipotence. And I think that, even if one handles these distinctions with the most meticulous care, it still seems very difficult to see how Van Inwagen's particular efforts here make the falsity of determinism look feasible. Remember that this example will secure Van Inwagen's objective if and only if it demonstrates the falsity of both the linguistic and the metaphysical regularity theses about causation. The falsification of the linguistic thesis alone won't do, since this is consistent with the truth of the metaphysical thesis, which simply is the truth of the claim that causes do, in fact, determine their effects. Let us look first at the linguistic thesis. This is the claim that the whole concept of causation without determination is incoherent. Van Inwagen's suggested falsifying of this seems to take the form: "God reveals to us that causation without determination is a coherent concept, therefore causation without determination is a coherent concept". This strikes me however, as completely untenable. This is because God has no more facility than you or I to make what is incoherent coherent. As has been well-documented, He cannot make a square circle and, generally, if He is going to reveal to us what is coherent, it must be coherent in the first place. God can only reveal to us that causation without determination is a coherent concept, if causation without determination already is a coherent concept. He just could not, for instance, reveal that the concept

of a square circle is a coherent one, because, before God even thinks about speaking, the concept of a square circle is incoherent. So, God's word appears to offer no assistance whatever, with the task of dismantling the linguistic regularity thesis about causation. What about the metaphysical thesis then? Let us assume for the moment (as I, in fact, believe is the case) that the linguistic thesis is false and that, therefore, it is a possibility that the metaphysical thesis is false. Again, I do not see that the invocation of the word of God helps the indeterminist cause at all. We can even assume that God has pretty much a free hand here, i.e. we can assume that He has all logical possibilities laid out in front of Him (of which causation without determination is one), and he simply picks whichever one He fancies. Again however, He can only go on to reveal that determinism if false if He, first of all, makes determinism false. A "revelation" on God's part that determinism is false, can only be a genuine revelation, can only be true, if determinism is already false (whether made so by Him or by some other means entirely). This is precisely what Van Inwagen and ourselves are hoping to discover, but I can only assume that the reason Van Inwagen introduces the revelations of God into things is because he is misled into thinking that God can "reveal" anything whatever, that God's word is a guarantee of truth. If God does, in fact, "reveal" only what is true, then He will only reveal to us that determinism is false if determinism is false. This, I repeat, is what is at issue. We may even allow, on the contrary, that God enjoys a little joke every now and then, and may just be "revealing" a falsehood to us. We would therefore have to be wary; perhaps, despite what God says, it may just be the case that the glass did have to shatter under the conditions in question. Again, we would have to investigate this, to try to establish whether God is telling the truth or not. I hope that I have made it clear why this effort by Van Inwagen strikes me as question-begging, and why the "word" of God doesn't really help us. And indeed, whilst I will repeat that (as I will expand on soon) I am sympathetic to Van Inwagen's contention that causation doesn't entail determination, not only does this example fail, in my view, to support it, but I fear that it is much worse than this, even. Indeed, it seems to me that Van

Inwagen's formulation here, not only leaves a whole load of questions uncomfortably hanging, but in fact makes the Humean assumption in question look reasonable. Not only that, but (crucially, I think) one of the problems I find with this example, will turn out, really, to be the same kind of problem which generates the ultimate misgiving I will have about the conclusions of Van Inwagen with regard to free action.

What worries me, at bottom, is this: suppose that there is (as Van Inwagen seems fairly sure there is) a possible world having exactly the same laws of nature as the actual world and having histories identical with that of the actual world in every detail up to the instant (in Van Inwagen's example) at which the stone came into contact with the glass, but in which the stone rebounded from the intact glass. How would we account for the happenings in this (possible) world? What would, or could we say, happened, took place, in such a world at the point where the stone did not, in fact, shatter the glass, but rebounded instead? In the actual world, it seems very much like the history of the world, conjoined with the laws of nature, provide a complete explanation of what happened, make it reasonable to entertain nothing <u>other</u> than the shattering of the glass, from the moment the stone makes impact with it. However, according to Van Inwagen, it is doubtful whether this is strictly true, since there is probably a possible world, with identical history and laws of nature, where the stone rebounds from the glass. As I've suggested though, what actually happens in such a world? Do the laws of nature suddenly take a back seat, "decide" to have a break from operations? The point is that, it seemed very much as though we had a complete explanation for things as they happened in the actual world; there seemed, as far as we were aware, to be no explanatory "gap" whatever. As I said, the history of the world conjoined with the laws of nature seemed to close the door on the possibility of any happening other than the shattering of the glass, once the stone made contact. However, Van Inwagen's suggestion seems to be that this was mistaken all along, that there was indeed an explanatory gap, that the history of the world plus the laws of nature still allowed the possibility that the stone

would rebound from the intact glass. And of course (it goes on), whatever it was that filled this gap in the actual world, there are possible worlds in which it may be filled differently, and the stone consequently rebound from the intact glass. This, however, is the notion which I find it very difficult to make any sense of. I ask again: what kind of things are going on in the "gap"? Magic? What is the source of our "illusion" that the history of the world plus the laws of nature provide a comprehensive explanation of the stone's shattering the glass? And why don't they? Van Inwagen seems to me to provide nothing by way of answer to these questions and (whilst I am ultimately sympathetic to the anti-Humean position) I must therefore conclude that his attempt here to discredit the supposed synonymy of causation and determination is a failure. Indeed, I feel it worth saying at this point that one of the reasons I will be left dissatisfied with Van Inwagen's enterprise, is a similar gap in explanation (i.e. in the case of free action) between intentional state and action. In the light of what we have just seen Van Inwagen say, the intolerable "gap" (to which I will come) is little of a surprise. This, I hope, will be borne out as I proceed.

It is, I think, worthwhile to pause here to re-establish exactly where we are, and what we've been trying to do. Despite the nature of the territory I have just been going over, it is essential to bear in mind that what we are really concerned with here is action, and ultimately, the question of free action. What we are presently examining is the possibility of a non-deterministic theory of human action, i.e. which will ground real <u>action</u> (and ultimately, which will ground <u>free</u> acts). One means to opening up this possibility was to see if we could dismantle the Humean assumption that causation entails determination, and then go on to see if we could find a set of undetermined causes which could ground real action (e.g. these could be undetermined causes in a natural part of a human being). One attempt at dismantling this Humean premise, the one suggested by Van Inwagen which we have just documented, looked to be a failure. However, this does not mean, of course, that no attempt at this task can succeed. Indeed, all that really matters

here is that this Humean premise is seen to be wanting when applied to cases of human action, i.e. if it can be demonstrated that human action admits of undetermined causes, then that it is all that will matter in this regard; we needn't worry whether it is true outside of cases of human action, and it is perhaps significant that the (doomed) example of Van Inwagen's which I have just cited does not concern a case of human action. As I've indicated already, it is through focusing specifically on human action itself that John Searle attempts to show that the Humean premise is not universally true, that in the case of human action at least, (in its linguistic version) it seems to fail. As I've also indicated, in so doing he also provides a theory of action which is, in ways, similar to the (Davidsonian) one Van Inwagen favours, and also clarifies modal questions concerning the relationship between intention and action. All of these issues tie together in what is in parts a highly penetrating offering of Searle's, but before setting down his efforts properly, I think it a good idea to re-sketch a brief outline of the significant points of impact of Davidson's theory of action.

The cental idea is very simple: an act is caused by the very factors its agent would adduce if asked to give his reason for performing it, to wit, by the agent's desire that a certain state of affairs should be realised and his belief that an act of that type was the best - or, at least, an unsurpassed - means to realising that state of affairs. Suppose, for example, that a certain man has raised his hand; and suppose that when he is asked why he raised his hand, he says that he did this because he wanted to vote for the measure before the meeting and believed that the way to vote for this measure was to raise his hand. If what he says about the reasons for his act is correct, then, according to Davidson, the cause of his act is his desire to vote coupled with his belief that he could realise his desire by raising his hand. Davidson, remember, holds what I've said is a straightforwardly deterministic version of this theory which I've so far argued to be incompatible with the idea of free action (despite Davidson's aforementioned dismissiveness towards incompatibilism). Van Inwagen, as I said, is

sympathetic to an indeterministic version of it but, as I've just been arguing, I am unimpressed with his efforts at completing the necessary step to this end, of refuting the Humean premise, which is accepted by Davidson. When I was discussing Davidson specifically, near to the beginning, I expressed sympathy myself with Davidson's theory of action (as it is outlined here) and in Intentionality, Searle argues, not only that this kind of theory is true but, because of the nature of intentionality, it actually has to be true. As I've said, he also provides clarity on the nature of the relationship between intention and action. As I've also suggested, I think he fills in some of the gaps in Van Inwagen's effort, though perhaps not all of them. Let us look at Searle's efforts in "Intentional Causation".[9]

I laid out, a bit earlier, Searle's recounting of the three features of the orthodox view of causation, one of which was the assumption I have just been talking about regarding causation and determinism. As I've already said, it is certain sorts of very ordinary causal explanations having to do with human mental states, experiences and actions in particular, which Searle wants to say do not sit very comfortably with the orthodox account of causation. To take an example of Searle's, suppose I am thirsty and I take a drink of water. If someone asks me why I took a drink of water, I know the answer without any further observation: I was thirsty. Furthermore, in this sort of case it seems that I know the truth of the counterfactual without any further observations or any appeal to general laws. I know that if I hadn't been thirsty right there and then I would not have taken that very drink of water. Now when I claim to know the truth of a causal explanation and a causal counterfactual of this sort, is it because I know that there is a universal law correlating 'events' of the first type, my being 'thirsty', with events of the second type, my drinking water under some description or other? And when I said that my being thirsty caused me to drink the water, was it part of what I meant that there is a universal law? Am I committed to the existence of a law in virtue of the very meaning of the words I uttered? Part of the difficulty, as Searle notes, in giving affirmative answers to these

questions is that I am much more confident of the truth of my original causal statement and the corresponding causal counterfactual than I am about the existence of any universal regularities that would cover the case. Searle suggests that maybe there are physical laws at the neurophysiological or even molecular level that would describe the case (and, of course, this is precisely what Davidson's view of the matter is), but I certainly do not know for a fact there are such laws, much less what they might be, and I do not in making my original causal claim commit myself to the view that there are such laws. (Although I think Searle is correct about this, part of his analysis of this example helps generate the ultimate worry I will express about the efforts of both Searle and Van Inwagen. I will be returning eventually to this.) The claim that there are causal laws corresponding to the events in question is not a logical consequence of the singular causal statement. It is logically consistent to insist on the truth of the singular causal explanation and yet deny a belief in corresponding causal laws. As Searle says, I know, for example, what made me take the drink of water: I was thirsty, but when I say that I am not committed to the existence of any causal law, even if in fact I believe there are such laws. (It should be clear that this amounts to a denial of the linguistic thesis, but not of the metaphysical; the idea of undetermined causes of action and therefore the non-deterministic version of Davidson's theory which Van Inwagen favours demands, of course, that the metaphysical thesis also be false. As I will reiterate as I go on, Searle in fact, believes that the metaphysical thesis is likely to be true. This would, of course, have bad consequences for free will. Nevertheless, Searle's work on intentional causation does give compelling reason to believe that some version, deterministic or non-deterministic, of Davidson's theory is true, as I will document soon.) Searle points out that the Humean premise, in its linguistic version, makes an extraordinary, strong and unsupported claim, namely that any statement, such as

My thirst caused by drinking entails a statement of the form

There is some law L such that there is some description ϕ of my thirst and some description ψ of my drinking, and L asserts a universal correlation of events of type ϕ and events of type ψ.

And, of course, this supposition is hardly intuitively plausible. The linguistic thesis seems simply false; I think that Searle provides compelling reasons here for not accepting it. This closes the door, it seems, on at least one possible set of reasons for accepting the metaphysical thesis, since, of course (as has been said already) the linguistic thesis entails the metaphysical thesis. However, the metaphysical thesis may yet be true and (as I've also just said) this is really what ultimately matters for free action; even the truth of the Davidsonian theory of action which I've outlined (which Van Inwagen, Searle and myself all favour) does nothing, for free will, if both incompatibilism is true (as I believe it is) and the metaphysical thesis about causation is true (as Davidson and Searle both seem to believe). To say that the metaphysical thesis about causation is true is to say that determinism is true, that causes <u>in fact</u> determine their effects. I have already looked at Van Inwagen's attempt at refuting this thesis, in the shape of his "stone breaking glass" example, and found it wanting. The metaphysical thesis may yet be true (on non-linguistic grounds), and we must hold judgement on it for the moment. As I've said, I think it more than useful, for the moment, to look at Searle's comments on why a Davidsonian account of action (deterministic or otherwise) must be true. We must beware of losing sight of the main issues here; the possibility under consideration, remember, is that of finding a set of causes of human "action" which satisfy two conditions:

(i) they are undetermined,

(ii) they are such as to ground a belief that "actions" in question are real actions (and later, real <u>free</u> actions).

As I've said, the question of (i) depends on the truth or otherwise of the metaphysical thesis about causation, something on which we

must wait. (ii) on the other hand, turns upon the cause satisfying certain other kinds of conditions e.g. (as Van Inwagen has noted) it could be that the occurrence of the causes in question in a natural part of a human being would mean the satisfaction of (ii), and this in turn would be supported by the truth of the Davidsonian theory of action. The reason why, therefore, we are ultimately interested in adducing reasons for accepting the Davidsonian theory, is that it would seem to help establish the truth of (ii), leaving us then to concern ourselves only with (i). Let us look now then at Searle's valuable insights, insights which I think make it clear that some version of Davidson's theory must be true.

We can best begin this, I think, by considering the issue (and Searle's remarks on it) which I have mentioned at several junctures, and which is entailed by the third premise (which I set down) of Searle's statement of the orthodox view of causation. This concerns the question of the 'logical' or 'contingent' nature of the relationship between intentional states and action. Searle presents a very persuasive case that there is a logical relation of a sort (much weaker than the entailment relation between statements) between cause and effect in the case of action. That is, the cause itself quite independently of any description, is logically related to the effect itself, quite independently of any description. How is this possible? This is because, in the case of action, the cause is a presentation or representation of the effect e.g. thirst, regardless of how described, contains a desire to drink, and that desire has, as conditions of satisfaction, that one drinks; an intention in action to raise one's arm, regardless of how described, has as part of its conditions of satisfaction, that one's arm go up. The reason that there is a logical or internal relation between the description of the cause and the description of the effect in the cases mentioned is that in every case there is a logical or internal relation between the cause and effect themselves, since in both cases there is an Intentional content that is causally related to its conditions of satisfaction. (There may, throughout the Searle exposition, be a small measure of Searle's baggage, which I can think can generally be assumed.) The specification of cause and

effect under those causally relevant aspects involving Intentionality will give us logically related descriptions of cause and effect precisely because the cause and effect themselves are logically related - not logically related by entailment, but rather by Intentional content and conditions of satisfaction. Indeed, as Searle specifies, there is an identifiable formal structure of the phenomenon of Intentional causation in the case of action: there is a self-referential Intentional state or event, and the form of the self-reference is that it is part of the content of the Intentional state or event that its conditions of satisfaction (in the sense of requirement) require that it cause the rest of its conditions of satisfaction (in the sense of thing required) e.g. if I raise my arm, then my intention in action has as its conditions of satisfaction that that very intention must cause my arm to go up. That is, in the case of action, at least one term is an Intentional state or event and that state or event causes its conditions of satisfaction. More precisely, when an Intentional state x causes an action y then

(1) y is (or is part of) the conditions of satisfaction of x, and
(2) the Intentional content of x is a causally relevant aspect under which it causes y.

Indeed, where any causal theory of action is concerned, it is important to note that, in spite of orthodox ideas about causation, we directly experience causation in many cases when we make something happen. When, for instance (to take Searle's example again) I raise my arm, part of the content of my experience is that this experience is what makes my arm go up. In such a case, we directly experience the causal relation, the relation of one thing making something else happen. As Searle notes, I don't need a covering law to tell me that when I raised my arm I caused my arm to go up, because when I raised my arm, I directly experienced the causing: I did not observe two events, the experience of acting and the movement of the arm, rather part of the Intentional content of the experience of acting was that that very experience was making my

arm go up.

Despite the standard theory, Searle is indeed correct to say that every experience of acting is precisely an experience of causation. Whenever we act on the world, we have self-referential intentional states, and the relationship of causation is part of the content (not the object) of these experiences. Searle suggests, in fact, that what has been happening is that the Humeans' have been looking in the wrong place. That is, they sought causation (force, power, efficacy, etc.) as the object of perceptual experience and failed to find it, whilst Searle is suggesting that it was there all along as part of the content of both perceptual experiences and experiences of acting (whilst it needn't concern us here, Searle makes out an analogous and similarly penetrating case concerning perception). Again, when I raise my arm I don't see causation or raise causation, I just raise my arm. The movement is not part of the content of the experience, rather it is the object of the relevant experience. But causation is part of the content of the experience of that object. Searle's point is that it is not in the observation of actions that we become aware of causation, it is in the performance of actions, for part of the Intentional content of the experience of acting when I perform intentional actions is that this experience causes the bodily movement. Part of the actual phenomena of the action is the experience of causation. The experience itself does the causing and, where successful, it causes what it is directed at. The causal nexus is internal to the experience and not its object.

As I've been saying, I think that the foregoing considerations adduced by Searle are penetrating and persuasive, and reveal why a theory of action along the lines of that forwarded by Davidson and favoured by Van Inwagen must, at bottom, be true. Searle is aware that there are some possible objections to his account, but he goes on to demonstrate also, I think, that these can be satisfactorily negotiated, and I don't feel it essential to say much more.

So, where exactly does all of this get us to, and (just as

important), where doesn't it get us to? As I've said, Searle's offerings seem to make clear why a (Davidsonian) causal analysis of action cannot but be correct. However, where does this get us regarding the crucial questions of determinism, indeterminism, action, and free action? We can bring to mind also at this point that it seems to have been clearly established, that so far as the concept of causation is concerned, a non-deterministic account (despite Davidson's own position) of Davidson's theory of action is perfectly feasible (the linguistic "regularity" thesis about causation is false). This would mean that, if we adopt the stance Van Inwagen seems to do, of having an "action" being caused by an intentional state sufficient for its being a genuine action (waiving the question of free action for the moment), then an undetermined act would undoubtedly seem to be a conceptual possibility. Conceptually, an "action" may be caused by the intentional state of its doer (making it a real action), and may yet be undetermined. Fair enough. On the conceptual level, an undetermined act is a possibility. So if the second strand of the Mind argument is only that the concept of an undetermined act is incoherent, then it seems to fail. However, I hope it is clear by now why this, of itself, doesn't have a great deal of import. What is still unresolved is the crucial question of whether the metaphysical regularity thesis about causation is true; we have established that the idea of an undetermined act is coherent, but even if a Davidsonian account of action is correct, whether there are in fact any undetermined acts will turn upon whether the metaphysical regularity thesis is true: if it is true (as Davidson thinks it is, this indeed being entailed by his endorsement of the linguistic thesis), then there are, in fact, no undetermined acts; if it is false, then there are, in fact, undetermined acts. And, of course, it should be obvious from now I have been arguing to far that this brings in its train monumental consequences for the idea of free action, in my view, i.e. if undetermined acts are, in fact, impossible, then free will is impossible, since determinism would then be true and (I have been arguing) incompatibilism is true. If undetermined acts, are in fact, possible, then the problem for free will becomes that which the third strand of the Mind argument pinpoints, i.e. how an

undetermined act can be a free act.

Indeed, it seems to me that this is what is really the root problem with the Van Inwagen "breaking glass" example - he confuses the linguistic regularity thesis with the metaphysical regularity thesis. If it were merely his business to reveal to us that God could reveal to us, given our concept of causation, that determinism is false (i.e. could say "determinism is false" without incoherence), then he undoubtedly succeeeds. God could indeed tell us without incoherence that determinism is false. Such a suggestion is no over-investment in God's omnipotence, is not like saying that God could create a square circle. However, Van Inwagen both wants and needs to be claiming more than this, i.e. he needs, for his own purposes, to be saying that determinism is, in fact, false, that an undetermined act is not only a conceptual possibility but is also a metaphysical possibility, is compatible with the history of the world and the laws of nature taken together. And, as I've reiterated, conceptual possibility is hardly any guarantee of metaphysical possibility; Van Inwagen seems to speak as though something which doesn't follow does follow, i.e. it doesn't follow from the fact that God can say without incoherence that determinism is false that God utters a truth when He says "Determinism is false". In order for the latter to be the case, we need something more than conceptual feasibility: we need determinisn to, in fact, be false. And whilst it may be up to God whether determinism is false, (as I've said) the metaphysical fact or otherwise of the falsity of determinism is something independent of God's (coherent) statement of the falsity of determinism. God would first have to create a world where determinism is false, before He could state truly that determinism is false; though regardless of whether or not determinism is, in fact, false, God could indeed sit all day long in heaven and shout down perfectly coherently "Determinism is false!" As I also said, whilst I am persuaded that the concept of the falsity of determinism in a universe of cause and effect is a coherent one, in so far as Van Inwagen's "stone breaking glass" example is an attempt at demonstrating that determinism is, in fact false, I think it fails. More is needed, I believe.

As I said, the truth of the metaphysical thesis amount to the truth of determinism, and (in my view) the absence of free will. If the Davidsonian theory of action is sound (as I believe there are compelling reasons for holding), then the falsity of the metaphysical thesis means that there are undetermined acts, leaving then the question of if and how we can say that such acts are free. What of the metaphysical thesis? Davidson obviously holds it and, as I said, Searle thinks it likely to be true. From "Intentional Causation":

> ...these causal claims [singluar causal claims made with respect to four examples of intentional causation cited by Searle] do not commit me to the existence of any relevant causal laws. I might additionally, and as a matter of fact I do indeed, believe that there probably are causal laws corresponding to these four types of events but that is not what I meant...[10]

Two of these examples, of course, of intentional causation, are the two I have already mentioned, my thirst causing my drinking and my intention to raise my arm causing the raising of my arm (the other two concern perception). We can see that Searle thinks it likely, that there are, as a matter of fact laws at some level or other, relating cause and effect (à la Davidson). Therefore, according to Searle, determinism is, in fact, true (though not entailed by these singular causal claims); or probably true, at any rate. If my incompatibilist claims so far are sound, then this would entail that I cannot act other than how I do act. I could not, when I took the drink, have done otherwise, nor when I raised my arm, could I have refrained from raising my arm. This is what makes Searle's assurance that "it's up to me", whether or not I take a drink of water when I'm thirsty, puzzling: if there are laws at some or other bedrock level, which determine whether I take a drink of water, then in what substantive sense is it "up to me" whether I take a drink of water? It certainly would not be up to me in the sense which matters for free will. When dealing specifically with Davidson as a language strata theorist, I argued that Davidson's

position was incompatible with the existence of free will; what makes it so, I believe is the causal theory of action conjoined with the metaphysical regularity thesis about causation. And Searle subscribes, largely, to both of these, and therefore leaves us no better off with regard to free will. As I've said, Van Inwagen believes the metaphysical thesis to be false (despite his weak effort, in my view, at demonstrating it); as I also said, he leans in favour of the (Davidsonian) causal analysis of action.

Let me say at this point that as I already indicated when speaking of Van Inwagen's illustrative example, I am myself sceptical as to whether the metaphysical regularity thesis is false with regard to cases not involving human action; as I said then, the "gap" in the Van Inwagen kind of example has to be accounted for if the metaphysical thesis (with regard to non-action cases, at least) is to be false. I'm not entirely sure what could fill it. Let us assume for the moment, however, that it is false where cases of human action (despite Davidson and Searle) are concerned, i.e. not only have we real acts (because, as I said, we have "acts" which are caused by the subject's desires, beliefs, etc.), but we have undetermined acts. This is what the causal analysis of action along with the falsity of the metaphysical regularity thesis about causation would give us: actual undetermined acts, and not the mere conceptual feasibility of it (as the falsity of the linguistic thesis alone conjoined with the causal analysis of action would give us). We have actual, full-blown, real, undetermined acts. This would mean that the second strand of the Mind argument would have been negotiated. Welcome the third strand now, however. How could we say that an act, such as I have just described it, is a free act? It shouldn't be difficult to anticipate by now the precise import of something I said before, namely that the explanatory "gap" within Van Inwagen's "stone breaking glass" example (as he documented it) would be echoed and unsatisfactorily negotiated in the parallel attempt at setting down conditions of undetermined, free action. Van Inwagen, remember, isn't especially hospitable to the idea of agent-causation, and so (in parallel with his "shattered glass" example) he rests content, it seems, with saying that, whilst

we may not really know how this is the case, an intentional state which, on one occasion, caused a particular act may on another occasion (in another possible world) not have caused the same act even when this particular intentional state was the only thing which was causally relevant. This idea, with no further explanation, strikes me as curious indeed. The addition of extra causally relevant factors, on the "other occasion" would, of course, make all the difference, but they aren't there. Again, we have this gap: the world has a given history, and a given set of laws of nature, and I have a particular intentional state, this being the only thing which is causally relevant to what I do. Yet, somehow or other, just at this point, and not before it, causation may suddenly come up against a brick wall. There is (the story goes) something about this point which provides the option of causation "not working". Whether it does, in fact, work or not is (according to Van Inwagen) a complete mystery. For instance, on one occasion my thirst may cause me to take a drink of water; however, on another occasion, with an identical set of laws of nature and my thirst again being the only causally relevant factor, I might not take a drink of water. What actually happens at this crucial point, be it a millionth of a second or whatever, where causation either "works" or it doesn't? How might it not work? Van Inwagen states this odd position by saying that he must reject the following proposition:

> If an agent's act was caused but not determined by his prior inner state, and if nothing besides that inner state was causally relevant to the agent's act, then that agent had no choice about whether that inner state was followed by that act.[10]

He admits to finding it "puzzling" that this proposition should be false. I, however, find it much worse than that. I, really, find it intolerable. Not only that, but in light of much of what I have said already, something even more fundamental may, perhaps, be apparent concerning Van Inwagen's effort. That is, if this is really the best Van Inwagen has to offer, then he seems to locate freedom between the intention and the action intended, and not (as I

think it reasonble to do) at a step prior to the intention or act of will. I have made this point already at several junctures, that the significant questions for free will must include at least questions about the conditions of intention, choice, desire, belief, act of will, etc., and not merely questions about the path from intention (or whatever) to action. So, even if we could make sense of the supposed falsity of the proposition just stated, it's not clear that it would be of very much help for free will.

Van Inwagen claims that he finds all the alternatives inconveivable. However, perhaps he has not done justice to all possible alternatives. This is where some kind of agent-causation may be on hand to help. What if we added in to the causally relevant factors, the agent himself? On one occasion when I'm thirsty, I take a drink of water; on another such occasion I don't. It seems (prima facie) that, on both occasions, the only causally relevant factor was my thirst (coupled, strictly I suppose, with the belief that to take a drink of water would satisfy my thirst). But perhaps I am a causally relevant factor. Perhaps, when faced with thirst and a glass of water in front of me, I either cause my arm to reach out and take the glass up to my mouth, or I don't. Perhaps of causal relevance are not only intentional states, but myself also the bearer of these intentional states. Such an idea could perhaps fill in the gap which I find unacceptable in Van Inwagen's indeterministic, causal effort. Agent-causation may seem, prima facie, somewhat mysterious, but it can hardly be any more so than what we have just seen Van Inwagen suggest. Van Inwagen seems to speak as though agent-causation and his (Davidsonian) causal analysis are mutually exclusive, but I do not see that they are. On the contrary, (as I've just said), perhaps the addition of the agent himself to the existing stock of causally relevant factors may provide (as I suggested) the least repugnant analysis. It will be useful, I think, in helping us towards a positive theory, to look at some work by Hugh McCann, his article "Volition and Basic Action"[11].

Before doing this, I expect it will be no surprise that I don't take the third strand of the Mind argument to be any better a

positive compatibilist argument than the first, or second strand. It does nothing by way of counter-evidencing Van Inwagen's First Formal Argument. It can, at best, be an argument for the incompatibility of indeterminism and free action. And what I am presently trying to discover is a way to avoid just this conclusion.

Notes

1. In Physics, 256a.

2. In "Freedom and Action" (Freedom and Determinism, ed. Lehrer).

3. In "Descartes' Evil Genius", Philosophical Review, 1949 and "Descartes Scepticism of the Senses", Mind, 1945.

4. In G.E. Moore, "Proof of an External World", Philosophical Papers.

5. In "Actions, Reasons and Causes", (Essays on Actions & Events).

6. In Essay on Free Will, p.138.

7. In Essay on Free Will, p.140.

8. In Intentionality, Chapter IV.

9. In Intentionality, p. 120.

10. In Essay on Free Will, p.149

11. In "Volition and Basic Action", Philosophical Review, 1974.

11. TOWARDS A POSITIVE THEORY?

It is worthwhile, I think, before looking closely at McCann, to both state the parameters of McCann's exercise and remind ourselves exactly where we are and what we are trying to establish. I have, I hope, presented sound arguments for incompatibilism; if determinism is true, then there is no free will. The problem now being confronted is how there can be free will if determinism is true; so far, the notion that we ever act freely whilst determinism is false seems almost as intractable as the notion that we ever act freely whilst determinism is true (i.e. as intractable as I have argued the latter to be). I have just said that, whilst favouring Davidson's causal analysis of action and Searle's penetrating suggestions in this area, I find Van Inwagen's final position on how free will is compatible with indeterminism unacceptable. Remember that I did say, however, that I am in sympathy with the overall spirit of Van Inwagen's project. That is, I believe that he is correct to assume that any possible solution there is, is grounded upon the realisation that causation without determination is a coherent possibility (although, as I said, the falsity of determinism requires the actuality of a set of undetermined causes, the falsity of both the linguistic and metaphysical 'regularity' theses about causation). Van Inwagen, as we saw, has (like some others) suggested that the notion of an undetermined cause occurring in a natural part of a human being, could well bear fruit and, despite Van Inwagen's scepticism about agent-causation (or 'immanent causation', as it is sometimes called), a notion going back to Aquinas[1] and having modern-day back-up from Chisholm[2] and Taylor[3], I am presently attempting to discover if this idea can be exploited in order to help the spirit of Van Inwagen's enterprise become flesh. Near to the end of "Volition and Basic Action", McCann expresses recognition of the philosophical context which, through the crucial notion of the essay, he places himself in:

> If terminology is needed to distinguish those actions which have results from those which do not, I would suggest the words 'transient' and 'immanent' respectively, which were used

in traditional philosophy to mark a distinction at least very close to this one.

So, this crucial notion McCann is exploring is that of actions which do not have results (volitions), which (as he acknowledges), is very like the old idea (which I've already said a bit about) of 'immanent causation'. Before discovering what consequences this idea could possibly have for the hope of free actions being compatible with indeterminism, it will be necessary to explain it in some measure, and I will do this in a moment. But I will make one or two preliminary remarks first.

McCann is not arguing specifically that volitions actually take place, but is providing a theoretical account of the advantages of postulating volition. And perhaps one of these advantages will be a plausible-looking account of how undetermined free action is possible through a change in a natural part of a human being; or, at any rate (to invoke a highly useful distinction), if we can turn what looks like a philosophical perplexity into a problem, then progress will have been achieved. I hope it is clear that nothing offered in this area by Aquinas, Chisholm, Taylor or McCann, can have the slightest import for free will if determinism is true, i.e. we are assuming here that determinism is false. In fact, near to the end of this chapter, I will cite what I think is an important passage from the concluding section of "Volition and Basic Action", where I think McCann places his efforts very usefully within the entire context of the problem of free will. Indeed, I will be wanting to urge, over the course of both this and the next chapter ("Is Determinism Likely to be True?") that the significance McCann affords thought in his analysis, may well be more fruitful than McCann himself is prepared to assert, as regards the whole question of whether determinism is likely to be true.

The important concepts in McCann's enterprise are causally basic (or just 'basic') actions and volitions. Volitions are causally basic actions. This requires expansion. A great many of our actions may be said to consist in bringing about a change.

Raising my arm is bringing it about that my arm goes up, and moving a finger is bringing about the finger's motion. In some cases the change brought about is external to the agent; killing Smith is bringing it about that Smith dies. When an action is one of bringing about a certain change, that change may be called the result of the action in question. Hence the result of raising my arm is that my arm goes up; that of killing Smith is that Smith dies (these examples are McCann's). Results are events which are necessary for those actions whose results they are. But they are never sufficient for those actions. An event appropriate to serve as the result of an action A might occur without A occuring at all. If Smith dies someone may have killed him but no one need have. An upward motion of my arm does not guarantee that I have raised it (there is also the 'deviant causal chains' possibility, the importance of which is clearly brought out by the points we have just seen Searle make regarding the precise causal relationship between intention and action). This lack of sufficiency is what made Wittgenstein's question about arm-raising possible. McCann calls the general problem of answering such questions the action-result problem: it is the problem of providing an account of how it is when events and processes qualify as results of human actions, they do so qualify. As I have just said, we saw Searle go some of the way to providing an answer, but it may be that there is yet more to be said. Indeed, Searle is himself inhospitable to the concept of agent-causation (hardly surprising in view of his leanings in favour of the metaphysical 'regularity' thesis about causation), and one objective here is to examine whether the idea may be due a more favourable reception.

The initial step is to see why the idea of a causally basic action is even necessary. As McCann says, besides results, actions can also have consequences. Like results, consequences are changes one brings about in performing an act. But unlike results, they are not essentially tied to the action. Rather, they are caused by it. Thus the action of moving one's finger might have it as a consequence that a gun fires. Since the connection between an action and its result is intrinsic, the result of an action A cannot

also be a consequence of A. Frequently, though, the result of A counts as a consequence of an action B of the same agent, which is in some sense "other" than A. McCann notes that this fact, when it occurs, generally provides a solution to the action-result problem for A. For instance, we can go back to McCann's example of killing, where the agent brings about results external to himself. Killing Smith (action 'A') might involve bringing it about that a bullet enters his body (result of B). The entry of the bullet could not be a result of the killing, but it is the result of an act of shooting Smith (action B). If Smith dies as a consequence (result of A), then the result of the act of killing Smith is caused by the act of shooting him. We would say in such a case that the agent killed Smith by shooting him. Note, however, that shooting Smith does not cause the agent's action of killing him: it causes only the result of that act. We can now move a step towards the idea of a causally basic action. The pattern of action at issue is that in which an action B causes the result of an action A of the same agent, but not A itself. In such cases, we can say that A is a causally nonbasic action, and that B is causally more basic than A. It should be clear that the causal pattern is especially useful for dealing with the action-result problem, since it allows us to explain how the result of the less basic action A came to occur at all, and explain it in terms of action on the part of the agent. Thus if killing Smith involves a causally more basic act of shooting him, we can appeal to the shooting to explain Smith's death. If we ask the Wittgensteinian question "when I kill Smith, what is left over if I subtract the fact that Smith dies?", the answer is "my shooting Smith". The explanation amounts to a description of how the agent brought Smith's death about, and bringing about this result is the action of killing Smith. Thus when an action A involves a causally more basic action, the fact that it does provides a solution to the action-result problem for A. The result of A qualifies as a result because it is brought about by performing the causally more basic action B.

So far, so good. But it should hardly be difficult to imagine where the problem is going to arise. Suppose B also has a result,

as shooting Smith has its result in the bullet entering his body. We are then obliged to ask how this event qualifies as a result. And if answering it involves pointing to an action that is causally still more basic, in this case firing a gun, then we are clearly on a path that must have an end. For if every action encountered in this type of analysis involves both a result and a causally more basic action, one would have to bring about an infinite series of further changes in order to bring about any change or set of changes at all. As McCann says, people cannot do this, but they perform actions with results all the time. Hence the analysis of such actions in terms of the causal pattern must eventually terminate in an action that does not involve such a sequence. Such actions are causally basic, and it looks as though if such an action has a result, that result does not occur as a consequence of a causally more basic action. But of course, can causally basic action have a result at all? Many recent philosophers have believed that the causally basic action in the case of killing Smith is one of moving a finger, a consequence of which is that the trigger of the gun is depressed. If so, we can use the causal pattern to analyse what the agent does down to this level. But this can't evade the action-result problem, for bodily actions like moving a finger always have results: that of moving a finger is that the finger moves. That of raising an arm is that the arm goes up. And if the appeal to causally more basic actions is ruled out, we must solve the action-result problem in these cases by different means. This is where the idea of volitions could be very useful. McCann makes what seems to me to be the highly persuasive suggestion that the causally basic actions are volitions, which are themselves a species of thought. Before going on to look at what McCann adduces in favour of this suggestion, let me repeat that, even if McCann's promptings are as persuasive as I think, there may well persist very dark areas with regard to how free will is possible. I have already said that we may eventually have to rest content, for the time being, with what is merely the least repugnant analysis, and also that if we reach a point where we have a clearer grasp of what form any acceptable solution would take, then something will have been achieved. Van Inwagen, remember, admits to a final puzzlement

(which I have said I find intolerable), and it may be that we can improve on that.

McCann is, like myself, aware that it is possible to question the strategy of treating thinking as an action, and I will later cite some suggestions of McCann in favour of this approach. Let us assume it for the present, however. The action-result problem does not seem to arise about thoughts. Unlike acts of moving a finger or flexing a muscle, thoughts do not seem to have results. McCann provides the example of the mental act of thinking of the number 1. Here there seems to be no event I bring about which is logically required for the act's occurrence, yet not sufficient for it, as the notion of my finger is in the case where I move it. Nor can my act of thinking of 1 be considered its own result. Results cannot be sufficient for the occurrence of the actions whose results they are, but thinking of 1 is sufficient for thinking of 1. It seems, then, that to think of the number 1 is not to bring about any result at all. And (if McCann is correct) the same would seem to apply to all thinking. If acts of thinking do not have results there can be no action-result problem about thinking. If there is no result to be distinguished from an action, there can be no question as to what makes it a result. If all this is sound, it follows that all acts of thinking must be causally basic (as I've already indicated, the whole notion of thinking as a causally basic activity will assume substantial importance as I come round to consider the whole question of whether determinism is likely to be true). The worth of this notion is that it makes acts of thinking especially suitable for providing a solution to the problem of the threatened regress, which arises if all actions that do have results are analysed in terms of the causal pattern. In what I think is a revealing, and indeed, rather ingenious example, McCann asks us to suppose that he is attached to a highly advanced electroencephalograph, which displays a characteristic pattern whenever he thinks of a one-digit number. An examiner asks him to produce the pattern, which he does by thinking of the number 1. Here the act of producing the pattern is causally nonbasic, since the occurrence of the pattern is caused by mental action. That action, however, is necessarily causally

basic. We could then say how the occurrence of the pattern qualifies as the result of the action of producing it, without analysing the producing of it in terms of any act about which the action-result problem arises. As I've already indicated, McCann's ultimate claim is that, since volition is thought, volition must be causally basic if it occurs at all. Supposing, therefore, that in the usual case where an action that has a result is performed, the causally basic action is one of volition, the action-result problem would be solved. If, for the moment, we waive the problem of whether determinism is true, there would seem to be a major advantage afforded by this proposed theory of volition. When a bodily change is brought about by performing a mental act, that act seems to serve as the means by which voluntary control over the change is exerted (I repeat my present waiving of the problem of determinism). It would thus constitute the exercise of agency through which the bodily change and its consequences come to qualify as results of actions. McCann suggests that in the case of volition, the mode or manner of the thought can best be described as the willing of the content. And it is this modality which makes it practical thought, rather than speculative, assertive, desiderative, or wishful. That is, volition is execution: to will the occurrence of a change is to enter upon the act of bringing it about. I hope that it is beginning to look plausible that McCann's suggestions supplement the efforts of Van Inwagen, Davidson and Searle, which I have already documented. It will be remembered, I expect, that I voiced disquiet with the "gap" in Van Inwagen's proposed analysis; I said that I found it intolerable that an agent may perform a particular act on one occasion, and on another occasion have the opportunity to refrain from performing the same act, when the set of causally relevant factors is identical (i.e. a la Davidson, some set of desires and beliefs). I am here floating a possible means of filling in the gap left in Van Inwagen's offering, and it is very worthwhile, I think, to quote outright at this point, an extract from McCann's essay, an extract which makes explicit that McCann is, like myself, aware of the role volition can play towards closing the (to me, intolerable) gap in the analysis of Van Inwagen:

> This account neatly fills the gap between the mere having of reasons and intentions, and the occurrence of changes that are results of actions. If it is correct, the connection between the two <u>is</u> through the agent's executive thought. This provides a role for intention and reason, as guides to conscious action...the account does not as far as it goes, do any violence to the agent's voluntary control in action over the changes that occur in his body. For it would be through the action of willing that the control is exercised...Actions with results would be fully intentional on this account, provided the result was willed, and was caused in the anticipated way.

I hope that some clarity has now been given with regard to my original suggestion that, as well as intentional states, it may well be fruitful to see whether the agent himself can reasonably be regarded as a causally relevant factor. I think that McCann has shown that volition, as a causally basic action, gives some sort of content to this notion. On one occasion, I take a drink of water when thirsty. On another occasion, in identical conditions, I don't take a drink of water. If McCann's analysis is sound, what is present on the former occasion, and absent on the latter, is a volition i.e. my willing that I take a drink of water. This, since it causally basic, is something that I do without having to do something else. McCann also, I think, provides a persuasive defence of his claim that volition is, indeed, genuine <u>action</u>, with the single exception that it lacks a result. Suppose a person engages in the act of willing that his arm rise. McCann is right, I think, to say that, besides constituting a conscious act he can know he performs, it seems perfectly natural to say this act is done by the agent <u>in order</u> to raise his arm. McCann's most persuasive consideration is, I think, the case of paralysis. That is, suppose the arm does not go up, as it surely wouldn't in this case. As McCann says, it hardly follows that the agent has failed to act intentionally, and here only the volitional act would be performed. McCann also notes that it is indeed hard to see how volition could fail to be intentional. Going back to McCann's shrewd example cited

already, it is just like thinking of the number 1, where this is done to produce a wave pattern. And similar examples can be used to show that a great deal, at least, of our thinking is intentional. Whilst I feel it essential not to beg any questions at this point regarding what is, after all, the central issues I am addressing in this thesis, namely questions of free will, determinism and moral responsibility, it is true that (as McCann notes) ordinary moral discourse is replete with cases wherein people are held responsible for thoughts, and a person would be held responsible for whether he engaged in willing if, for example, the act was meant to constitute a text for paralysis. Whilst I think that McCann is sometimes guilty, in "Volition and Basic Action", of doing the kind of thing I have just warned against, i.e. running together volition, intention and responsibility, thus begging the question in favour of free will, he is, I think, right to say that the kind of considerations just adduced certainly support the contention that volition is action (I will of course, come to consider the import of all this for free actions, soon). Whether it is free action is, of course, another matter and ordinary moral discourse, in holding people responsible for thoughts, seems to beg the question here. Assuming that moral responsibility does indeed require free will (I will conclude by considering and rejecting some opposition to this claim), then we can only be morally responsible for our thoughts if we have them freely, i.e. if we could do other than have them. However, for the moment, if McCann's suggestions are generally sound, then volition has all the features of the normal intentional action of raising one's arm or moving a finger, except that it lacks a result.

As I've already indicated, neither McCann nor myself believe the offerings of "Volition and Basic Action" to be any kind of last word, not, at any rate, where questions of free will, determinism, compatibilism, incompatibilism and responsibility are concerned. It is essential, of course, not to lose sight of the fact that it is these issues which I am concerned with, and any concern I have given to the theory of action will only be of any help in so far as it helps us with regard to issues within free will and determinism. So the question which must now be addressed is this: assuming that

McCann's theoretical defence of the notion of volition is plausible (as I believe it to be), exactly where does it get us vis a vis the issues with which I am concerned? There is a perfectly simple sense in which I believe the answer to be "nowhere at all". That is, despite what I've just noted as seemingly question-begging, tendencies on the part of McCann, McCann's offerings do absolutely nothing so far as I can see to evidence any of free will, determinism, compatibilism or incompatibilism. Remember that the reason I am looking at him at all is to explore a possible way of making the compatibility of free will and indeterminism theoretically plausible; yet, nothing he has said (even if it is all defensible) itself entails the falsity of determinism (and therefore, of course, according to my incompatibilist claims the possibility of free will). That is, it seems to me that volitions may take place and be, in the sense which he suggests, causally basic, and yet determinism be true, and hence free will false. There is no incompatibility, in so far as I can see, between McCann's suggestions, and (say) Van Inwagen's (fantasy) (M) hypothesis, or Ginet's H hypothesis. Volitions may be causally basic actions of mine, they may be actions of mine which lack a result, actions which I perform without doing anything else first, yet they may themselves be determined by Martians, by physiological-cum-environmental laws, or in some sense by the past taken together with the laws of nature. I may be thirsty, and may or may not take a drink of water; whether I do, in fact, take one may indeed turn on whether I will to do so. Yet this last factor, whether I do, in fact, will to take a drink of water, may itself be determined by Martians, physiological-cum-environmental laws, or in some way or other the past taken together with the laws of nature. That is, my willing may, strictly, be as much outwith my control as are the past and the laws of nature. It may be causally basic in that I don't to anything else in order to do it, but if (M) is true, then a Martian does something in order that I do it, and so the volition is in no way a free action of mine. Similarly, with the truth of any determinist hypothesis. In the important passage which I alluded to earlier, McCann both expresses awareness of this feature, and provides what I think is probably a crucial

contextualising of his theory, with regard to the problem of free will:

> But the great problems about action remain. They are, of course, the free will problem and the related problems of the nature of agency (of which the action-result problem is only a relatively insignificant part), and whether desires and beliefs cause action. I see no decisive solution to any of them in what I have said here, largely because volition has primarily to do with the execution of intention, whereas the free will problem is mainly concerned with its formation. But both the formation and execution of intention are functions of mental agency...so perhaps what has been said here will help lead us in a more fruitful direction than some that have been followed recently. For the most important thing about the free will problem is that it is a problem about the will.

As I hope to clarify and defend through the remainder of this chapter and part of the next one, McCann's remark that "both the formation and execution of intention are functions of mental agency" could be crucial, both in helping discover whether determinism is likely to be true, and (optimistically) in helping ground a positive theory of undetermined, free action. On the question of the former, I hope it has been borne out to some measure by my preceding comments that, if thoughts are causally basic actions, then we seem, in a sense, to be back at square one in asking "are thoughts determined?". I will be returning to this in the next chapter.

On the question of the latter, what if determinism is false? How could McCann's offerings help us see how free action is possible? As I've said, I think that McCann's model by no means removes all mystery from this area. However, I think it does help us to at least see precisely where the problem of free will could be located if determinism is false, and provides us with a plausible theoretical framework within which any solution (should it ever be found) would sit. Or to put it again as I did earlier, it helps us make the move from perplexity to problem. Remember that I expressed

dissatisfaction with the "gap" in the account favoured by Van Inwagen. I am now saying that McCann gives us some reason to believe that this gap could be filled by another causally relevant factor, i.e. the agent. And since the formation of intention is also a function of mental agency (as McCann pointed out), it also seems not implausible to regard the agent as a causally relevant factor in the formation of intention.

Van Inwagen suggested (remember) that it wasn't clear that an undetermined change in a natural part of a human being couldn't be the cause of a free action, and I am now suggesting that there may be some plausibility to the notion that the agent himself (assuming determinism to be false, of course) could be the cause of such a change (or, at least, one causally relevant factor). Let us go back again to an example of McCann's. I produce the pattern on the electroencephalograph by thinking of the number 1. The mental action is causally basic; I, it seems, cause the mental action without doing anything else first (or at least, I certainly seem to be one causally relevant factor), which in turn causes a brain-change, resulting in the pattern displayed. If determinism is false, and this model correct, then it would seem to be the case that (as Van Inwagen favours) an undetermined change can occur in a natural part of a human being in such a way that free action may be grounded, and that I am a causally relevant factor in bringing about this change (this feature itself grounding the idea that the action is free and helping fill the intolerable gap in Van Inwagen's suggested Davidsonian analysis , the gap which makes free action look no less odd than if determinism were true).

I say again that I feel no temptation to invest too much in this notion. I repeat that it seems to provide no more than a theoretical framework which gives the idea of undetermined, free action some substance. I am perfectly well aware that it does itself throw up very difficult, and for the time being possibly even intractable questions. I admit that there is still something mysterious-looking about my simply causing something without doing something else first. Indeed, I readily concede that the whole

picture seems to throw up a whole set of worries normally associated with Descartes' project. The idea of my causing myself to think of the number 1, for instance, and this in turn causing a physical modification, i.e. a brain-change, seems a paradigm case for anti-Cartesian objections. It seems that we must, straight off, be assuming the existence of a self with certain causal powers, and itself standing in an enclosed, protected, yet causally powerful relationship with its thoughts (a set of which are its volitions). I humbly accept that such problems seem enormous. I cannot say that I am confident about what the self is; I cannot say that I know how such a self could have the required relationship with its volitions and how the latter could have the causal efficacy which is required. I am not confident either about how such a self would ground action, free or otherwise. I cannot say that I know how thought could have the autonomy, primacy, basicness and causal efficacy which the model requires. However, it seems to me that the squaring of free action with indeterminism requires something like the combining of Davidson's causal analysis and the idea of agent-causation, such as I have documented. At least, it would leave us knowing precisely what the problems are; it seems to me then, that any squaring of free will with indeterminism must inevitably lead us into the most basic Mind-Body territory. If one thinks that the kind of problems which I have left hanging, and to which I know no answer, are an intolerable burden, then if one also accepts my incompatibilist claims, one will prefer to entertain that free will is impossible. Of course, it may yet be that determinism is true and (assuming incompatibilism) free will false, in which case the intractable-looking problems I have just cited will hardly matter here. This is something I will now address, i.e. whether determinism is likely to be true.

Notes

1. In St. Thomas Aquinas, Summa Theologica I, 54,2.

2. In "Freedom and Action" (Freedom and Determinism, ed. Lehrer).

3. In Action and Purpose.

12. IS DETERMINISM LIKELY TO BE TRUE?

As I will elucidate, I think that there are reasons for believing determinism to be false, and hence free will a possibility. However, before providing these reasons, I think it essential for me to say something about a whole tradition of thought in this area, a position (or set of positions) according to which there is either a particular kind of reason (which I will document) for believing that determinism is false, or that the issues of determinism and compatibilism are at bottom, irrelevant with respect to the question of moral responsibility. Although there are, as I've indicated, a few variants of this position, I think it reasonable to group them together, largely because of the common debt they owe to Kant. The arguments are all, I think, essentially Kantian in character. Proponents have included Wilfrid Sellars[1], Daniel Dennett[2] and, most famously, P.F. Strawson[3]. Indeed, Donald Davidson's views on several related issues are essentially the same in spirit, and (as I will also elucidate) Van Inwagen makes a similar move at one juncture.

What these contributions also have in common is that they provide what seem to me a set of defective arguments, a set of defective reasons for believing that determinism is false, or that its truth or falsity could be of no relevance regarding issues of free will and responsibility. In truth, it is perhaps very valuable to isolate Van Inwagen from the others I have mentioned here, since it is really only he who believes that determinism looks false, on the kind of grounds under consideration. For the time being, let me say that I do not think Van Inwagen's claims on this point are sound, and I will return to them properly. The others however, are really the class of philosopher who believe that the question of determinism simply could not be of relevance to the questions of free will, moral responsibility, and the reactive attitudes which seem to depend on them (I hesitate to call this position compatibilist, since it seems, indeed, to be itself grounded on the claim that the whole idea of a compatibilist problem is illusory; Van Inwagen is different, in this regard). According to this, it is

not possible for the reactive attitudes to be philosophically undermined in general by any belief about the universe or human action, including the belief in determinism. This extract from near to the end of Strawson's essay, expresses the essence of this view:

> Inside the general structure or web of human attitudes and feelings of which I have been speaking, there is endless room for modification, redirection, criticism and justification. But questions of justification are internal to it. The existence of the general framework of attitudes itself is something we are given with the fact of human society. As a whole, it neither calls for, nor permits, an external 'rational' justification.

As Thomas Nagel acknowledges[4], Strawson's claims here run parallel with those he makes about knowledge. According to this, justification and criticism make sense only within the system: justification of the system from outside is unnecessary, and therefore criticism from outside is impossible. I have already mentioned in connection with Harry G. Frankfurt, the transition from the subjective standpoint to its objective counterpart, and I believe that Nagel is right to say that Strawson's position is incorrect because there is no way of preventing the slide from internal to external criticism once we are capable of an external view. As Nagel says, it needs nothing more than the ordinary idea of responsibility. The problem of free will, like the problem of scepticism, does not arise because of a philosophically imposed demand for external justification of the entire system of ordinary judgements and attitudes. It arises because there is a continuity between familiar "internal" criticism of the reactive attitudes on the basis of specific facts, and philosophical criticisms on the basis of supposed general facts. When we consider the possibility that all human actions may be determined by events in the past and the laws of nature taken together (or for that matter, by Martians), it threatens to defuse our reactive attitudes as effectively as does the information that a particular action was caused by the effects of a drug - despite all the differences between the two

suppositions. No new standards have come into play here, since all we are doing is generalising familiar standards of criticism. We cease to resent what someone has done if we cease to see the alternatives as alternatives for him, something which would be the case if both determinism and incompatibilism were true. Nagel himself usefully fills out the epistemological parallel here. That is, the extremely general possibilities of error that the sceptic imagines, undermine confidence in all our beliefs in just the way that a more mundane particular possibility of error undermines confidence in a particular belief. The possibility of complete erosion by sceptical possibilities is built into our ordinary beliefs from the start: it is not created by the philosophical imposition of new standards of justification or certainty. On the contrary, new justifications seem to be required only in response to the threat of erosion from ordinary criticisms, sufficiently generalised. In the case of action, some of the externally imposed limitations and constraints are evident to us. If we discover others, not as obvious, then our reactive attitudes towards the affected action tend to be defused, for it seems no longer attributable in the required way to the person who must be the target of those attitudes. And, as Nagel says, the philosophical challenges to free will are nothing but radical extensions of this encroachment. If our actions are, in fact, determined by the past conjoined with the laws of nature, then our reactive attitudes would be entirely defused, incoherent. The push to objectivity opens up this possibility and, as Nagel says, the push to objectivity is a part of the framework of human life. And it could only be stopped from leading to sceptical results in areas of knowledge, action or anything else, if the external view of human life could be shown to be illegitimate - so that our questions had to stop before we got there.

Indeed, not only am I in agreement with Nagel's remarks here, but I have been persistently nonplussed at both the frequency with which Strawson's and related positions have been expressed, and the respect which they have generated. In fact, it strikes me as so clearly incorrect that it requires more a diagnosis than a refutation. What is lying at the bottom of it, I think, is

something fairly primitive, and perhaps not fully articulable. However, a glimmer of light can be shed on what it actually is, through these lines from Wordsworth:

> Ye dreamers, then
> Forgers of daring tales! we bless you then,
> Imposters, drivellers, dotards, as the ape
> Philosophy will call you...[5]

The idea in question seems to be that philosophical analysis, the drive to objectivity, is some kind of rival which has to be beaten off, in certain areas at least. I said at the beginning of this section that there were variants of this position, and in the cases I have seen, I have sometimes found it difficult to know precisely what position is being propounded, and indeed I have sometimes thought that the same essay slides between different positions (I include Sellars and Strawson here). That is, it sometimes appears (as Nagel describes Strawson), that what is being said is that the questions themselves which appear perfectly real to Nagel and myself, are empty. Other times it appears that what is being said is that questions of determinism and free will are real, but only within certain contexts, when we have a certain set of ("scientific") interests motivating us; they are inappropriate from the viewpoint of human action and moral responsibility. Sellars' position, in particular, often seems to be this one, and my suspicion is that it is this one which Davidson really entertains as well; it seems to be a case of our having two largely unrelated possible standpoints on our behaviour, i.e. (loosely) the scientific/objective and the intentional. We adopt whichever one we fancy according to our motivation; the neurophysiologist who is interested in prediction (say) will adopt the former and from this standpoint, notions of free will and responsibility seem empty, but when we want to entertain reactive attitudes, we can switch channel quite calmly to the latter, and away we go, with free will, responsibility, praise, blame and all the rest of it. Yet another variant is one which acknowledges the reality of the problem, that determinism is a genuine thesis, that its truth would leave us (when

we are doing philosophy, at any rate) puzzled about the source of our reactive attitudes, but that we ought simply to make a gallant stand against the enquiring impulse on behalf of other aspects of our make-up. This seems to me a particularly desperate manoeuvre, and I think it is exemplified here in this footnote to Strawson's "Freedom and Resentment":

> ...Quite apart from the issue of determinism might it not be said that we should be nearer to being purely rational creatures in proportion as our relation to others was in fact dominated by the objective attitude? I think this might be said: only it would have to be added once more, that if such a choice were possible, it would not necessarily be rational to choose to be more purely rational than we are.

As I hope I've made clear by now, I believe that every strand of this general position is defective. The kind of criticisms made by Nagel really cut across every one of them. Nor (as I indicated) am I impressed by Van Inwagen's suggestion that the very best possible reason for believing, not only in the falsity of determinism, but in the truth of the free will thesis, is the undoubted reality of moral responsibility. If it is agreed (as it is by Van Inwagen) that moral responsibility requires free will, then it seems to me intolerable that one should accept the reality of moral responsibility in the absence of independent evidence in favour of free will. We must have reason to believe that we have free will before we can entertain any thoughts of moral responsibility; it seems to me fatuous to argue the other way. Free will (as the considerable labours of Van Inwagen himself bear out) is far too difficult a thesis to be made parasitic upon the kind of persuasive points in favour of moral responsibility, which are adduced by Van Inwagen. The kind of Strawsonian points Van Inwagen documents are not only familiar enough, but generally true, I would expect. He is right to say that people believe in the reality of moral responsibility, that they behave as though moral responsibility were a given. He is clearly right to say that few people will react to an act of gratuitous injury deliberately done them by a human being

in the way that they would react if that same injury were caused by a bolt of lightning or a bough broken by the wind. When some person injures us - at least if we believe he knew what he was doing and that he could have helped doing it - we react in certain characteristically human ways: we blame, we remonstrate, we hate, we reflect on the futility of hate, we plan revenge, etc. As I say, this is all uncontroversial enough, but it does not seem to me to be genuine evidence that we have free will, and that we are, in fact, morally responsible. These familiarities do not of themselves counter-evidence the thesis that all our actions are determined by events in the past and the laws of nature taken together; they do not counter-evidence (for instance) Ginet's H hypothesis or, indeed, Van Inwagen's fantasy (M) hypothesis. Whether the kind of facts Van Inwagen cites about our behaviour and attitudes are rationally justified is itself contingent upon whether we have free will, and this itself must be dealt with on its own independent merits. As we have already seen Van Inwagen himself warn, the world may appear to be exactly the same, regardless of the actual truth of the respective determinist and free will theses. Van Inwagen here seems to forget his own directive; Nagel's riposte applies as much, I think, to what Van Inwagen says here, as it does to the comments of Strawson and Sellars. As I've already said at various junctures, the questions of free will and moral responsibility can only gain a very limited input from ordinary appearances; this is really another way of making Nagel's point that the slide from internal to external criticism cannot be avoided. This realisation is really of monumental significance. In order to gain evidence either way about determinism, and hence about the possibility of free will (and therefore, moral responsibility) we cannot, despite what we have just seen Van Inwagen say, fudge the following question: is there any reason to believe that human actions are uniquely determined by the past and the laws of nature taken together? No matter how much philosophers may try to evade it (as so many tend to), there is no way of escaping the relevance of hard data on this issue.

But, as Van Inwagen himself documents, the evidence gives us little reason to believe that determinism is true. It is almost a

banality, for instance, that the standard interpretation of quantum mechanics is indeterministic. And there is, as Van Inwagen also notes, little to commend the objection that quantum-mechanical indeterminacy applies only to unobservable submicroscopic events and not to the observable events of everyday life. Quantum mechanics is a physical theory, and is thus as legitimately applied to the behaviour of a glacier, an elephant, or (most importantly, of course), a human being as to the behaviour of a neutrino or a positron. Also, of course, the domain of the submicroscopic and the domain of the observable are not causally isolated from each other. Individual submicroscopic events can "trigger" observable events. Van Inwagen provides the example of the Geiger counter. The passage of a single alpha particle or electron or burst of high energy electromagnetic radiation through a tube of gas can sufficiently ionise the gas in the immediate vicinity of the path of the intruder to produce an effect that - after suitable amplification, which a Geiger counter is designed to provide - registers an audible click. Generally, what is of real importance is that actual matter, matter that obeys the rules of quantum mechanics is intrinsically incapable of carrying within itself the perfectly deterministic dispositions to future behavour that strict determinism requires. And, whilst neither Van Inwagen nor myself understand the issues to any serious extent, twentieth-century physics has not supported any attempt at embedding quantum mechanics within a more general deterministic theory. Most importantly, of course, no evidence has been collected to support the belief that human beings are determined. To reiterate, then, there would seem to be two ways in which scientific evidence could convince us we are determined; first, we might believe this, as Laplace did, on the basis of our most general physical theories (which apply to all physical systems and hence to us); secondly, we might believe this on the basis of the empirical study of man. But our most general physical theories are no longer deterministic. And the empirical study of man has a long way to go before it will be in a position to tell us anything about whether we are or are not determined.

These points are all, I think, very important, but what I think

is a very persuasive and generally neglected point against determinism being true, is also the one which I think provides reason to believe that we have free will. Remember that the falsity of determinism is itself no guarantee of free will; as I devoted some time to, making free will compatible with indeterminism is also a very difficult task. What I want to say here does not, of course, entail either the falsity of determinism or the truth of free will, but it does, I think, make both seem likely. I said earlier that I think there is some merit in McCann's idea of volition as thought, the latter being a class of actions which lack a result, and whilst (as I said) nothing offered by McCann is incompatible with determinism, what I want to say here is germane in that it is thought, or certain features of it, which both make determinism seem unlikely, and ground the likelihood of free will.

Let me deal first with determinism. This may seem to some somewhat weak, and indeed some may not even take it seriously as an argument at all, but it seems simply unlikely to me that all of the thoughts which humans are constitutionally capable of having are uniquely determined by the past and the laws of nature taken together. That there are, in principle, laws of nature accounting for every thought (given the past) which everyone has ever had seems to me to be a very difficult thesis to believe. At any rate, its truth would seem to me to demand, at the very least, an extension of the concept "law of nature", which is (at the moment, anyway) difficult to entertain. That is, the idea that laws of nature could be the kind of thing which, conjoined with facts about the past, would account for every thought I have ever had (and, given the causal role of thought, every action I have ever performed), just seems something there is little reason to believe. Even the most sophisticated and detailed laws of nature which we entertain seem to fall far short of this possibility. There are certain things I do not want to deny, namely that laws of nature are, in principle, always discoverable, and (probably more importantly) that a very large part of the intentional lives of all human beings is relatively simple, straightforward, primitive. I appeciate, on the latter issue, that we are by no means all of us, all of the time,

trying to find ways of solving Zeno's Paradoxes, wondering how judgements of taste are possible or indeed, whether there is reason to believe that either free will or determinism is true. Perhaps the idea of determinism with regard to the more primitive features of the mental lives of human beings is not so unlikely, but this itself does not seem to me to alter the fact that it seems unlikely to be true with regard to the most sophisticated operations of which humans are intellectually capable. For instance, I have had, here, some very precise and detailed things, as opposed to some different set of very precise and detailed things, to say on the issues of free will, determinism, compatibilism, incompatibilism, action and responsibility. I find it very difficult to entertain the idea that the past and the laws of nature taken together could account for why I have said one set of very precise and detailed things as opposed, indeed, to an enormous set (in principle) of different, equally detailed and precise things. Is it really plausible to suppose that the past and the laws of nature taken together uniquely determine that (for instance), I support incompatibilism, think that the history of compatibilism is a catalogue of cheap tricks, think that some of Van Inwagen's attempts at disparaging the metaphysical 'regularity' thesis about causation are weak, or (for that matter) that Wittgenstein changed his view on the Picture Theory of Meaning or that Nagel thinks realism about the external world is likely to be true? I repeat my awareness that these features are all, strictly, compatible with determinism. They just make it unlikely, I think. And, as I've already indicated, this has back-up from neurophysiology, the present state of which does not evidence that the brain is a strictly deterministic system. These kind of considerations make it no accident, I think, that we have not yet discovered even <u>one</u> of the "physiological-cum-environmental" laws which Ginet postulates, and which would exhaustively account for the actions of (in Ginet's example) the man giving the paper. Of course, if Ginet's H hypothesis is true, these laws would account for all human action. My suspicion is that there are no such laws, none waiting to be discovered, even by the most sophisticated technique. Again, it seems very unlikely to me that the fact that the man in Ginet's example says one very precise, detailed set of

things to his audience, as opposed to any number of other sets can be accounted for by the past and the laws of nature taken together. I see no reason to believe that the Intentional can be entirely subsumed under laws of nature, despite the obviously intimate links between the intentional and the brain. Of course, it could be the case. But I think it unlikely. Despite (I repeat) the obviously intimate links between mental operations and the brain's workings, there seems to me to be some reason to believe, with regard to some domain at least, in a primacy of the mental, a basicness of the mental, i.e. mental operations which need not refer us back to a more basic neurophysiological, and which cannot be explained by reference to the past and the laws of nature taken together. Although I am well aware of the difficulties with this notion, it is also, at bottom, as I've already indicated, the source of my belief that free will is likely to be true. I have already said that I believe there to be merit in McCann's defence of volition, that volition is thought and is causally basic. It is this idea, taken together with what I have just said regarding the primacy of at least some thought, which grounds my belief that it is likely we have free will. That is, two premises, (i) volition as thought, and (ii) primacy of (at least some) thought. This may seem a very weak defence of free will, and I am aware, obviously, of major difficulties. I have already stated these, really, when I spoke specifically about McCann himself, and, of course, the problems seem to have a recognisably Cartesian flavour about them. How thought could be primary in the sense in which I speak of, how I could just cause a change in my brain or body without doing something else first, are questions to which I cannot say I know the answer. Nor does anyone, I would think. What is clear is that, if there is anything to be said at all for my suggestions, the Free Will Problem must inevitably lead on to problems of the Self and, of course, Mind-Body. Both of these issues, of course, demand a separate piece of work. I have little to say on either, except to float a suggestion of Nagel's on the Mind/Body problem, to which I think there is merit:

> Can one thing have two distinct essential properties that are

not necessarily connected with each other? This seems possible if the two properties are different aspects of a single essence. For example, a tiger is essentially both a mammal and a carnivore, but those two properties are not always linked. They are linked in the case of the tiger because both are parts of the essential nature of the species - a particular type of mammal that can live only on certain kinds of food, and has other essential charactertistics as well...presumably something similar would have to be true if mental processes had physical properties. They couldn't be just slapped together. Both must be essential components of a more fundamental essence.[6]

What such a fundamental essense could possibly be, I have little idea. And, as I've said, I do not think that we can properly grasp how free will is possible until we know at least something about this fundamental essence. That is the point where things are left, I think. The perplexities have at least been transformed into problems. As I've said and tried to defend, I think it most reasonable to believe

(i) incompatibilism is true,
(ii) determinism is false,
(iii) free will is true.

There are still, as I've acknowledged, deep problems. The problem of free will won't go away.

As should be clear, I have also been assuming all along that moral responsibility requires free will, that one is only responsible if one could have done otherwise. According to this, of course, the truth of determinism (assuming incompatibilism) would kill off moral responsibility entirely. I do myself share this belief, but it has recently met with some opposition. According to a paper by Harry G. Frankfurt[7], one can be morally responsible even if one could not have done otherwise. I disagree with Frankfurt on this, and will conclude with a brief examination of his comments.

Notes

1. In "Philosophy and the Scientific Image of Man".

2. In "Mechanism and Responsibility".

3. In "Freedom and Resentment".

4. In The View From Nowhere, Ch. VII (4).

5. In The Prelude, Book V, (521-526).

6. In The View From Nowhere, Ch. III, (5).

7. In "The Principle of Alternate Possibilities", The Journal of Philosophy, 1969.

13. COULD MORAL RESPONSIBILITY NOT REQUIRE FREE WILL?

> Why is it supposed that movements cannot be actions? One main reason can be brought out by noticing that the same movement in one sense of those words, may occur in both the knee-jerk reflex, and in the action of moving one's lower leg...All that needs to be said about this is that there is a perfectly clear sense in which the same movement does not occur in the two cases...They are not identical, or instances of the same movement-type, in that, one does not derive from intention, and one does.[1]

It is this kind of realisation made by Honderich, an observation about event-identity, and the role which intention (or its absence) may play in this, which, at bottom, provides a rebut to the suggestions of Frankfurt, which I am about to document. Indeed, I think that the general form of the rebut is remarkably simple, and that Van Inwagen's defence of the belief that moral responsibility requires free will is somewhat laboured. I will therefore myself devote only a little time to this discussion, since I think it doesn't really require much more.

> A person is morally responsible for what he has done only if he could have done otherwise.

Frankfurt calls this the Principle of Alternate Possibilities (PAP), and attempts to construct counter-examples to it. I will cite Van Inwagen's own Frankfurt-style example. Suppose there is a man called Gunnar who has decided to shoot his colleague, Ridley. Suppose a third man, Cosser, very much desires that Gunnar shoot Ridley. Cosser is delighted with Gunnar's present intention to shoot Ridley, but he realises that people sometimes change their minds. So he devises a plan: if Gunnar should change his mind about shooting Ridley, Cosser will cause Gunnar to shoot Ridley. Van Inwagen asks us to suppose that Cosser is able directly to manipulate Gunnar's nervous system, and is thus able, in the fullest and strongest sense of the word, to cause Gunnar to act according to

his wishes. We suppose, moreover, that there is nothing Gunnar can do about Cosser's intentions or about the power Cosser has over his acts. It would seem therefore, that Gunnar has no choice about whether he shoots Ridley. If he does not change his mind he will shoot Ridley. If he does change his mind, he will shoot Ridley. Every possible future that is open to him is a future in which he shoots Ridley. Van Inwagen recognises that these futures are of two sorts: those in which he shoots Ridley without having been caused to do so by Cosser and those in which he shoots Ridley only because Cosser caused him to. And perhaps he has a choice about which of these sorts the actual future will belong to. But he has no choice about whether the actual future will belong to one of these two sorts, and hence no choice about whether he shoots Ridley. Yet we would hold Gunnar responsible if he shoots Ridley without having been caused to do by Cosser. Yet he could not have done other than shoot Ridley. Does this falsify PAP?

I hope that the Honderich extract has provided a clue to my reply to this notion. Strictly speaking, it is not the mere shooting of Ridley which Gunnar may be held responsible for; despite the fact that Gunnar will shoot Ridley anyway, the centrality of intention, or its absence, is crucial to the individuation of the event. That is, Gunnar's shooting Ridley as a result of being caused to do so by Cosser is a different event from his shooting Ridley as a result of his own intention to do so. In the former case, Gunnar clearly could not have done otherwise: he could not do other than shoot Ridley as the effect of Cosser's causing, and consequently (by PAP), Gunnar is not held responsible for shooting Ridley. In the latter case however, Gunnar could do other than shoot Ridley as a result of his intention to shoot Ridley (assuming free will, of course), and it is this precise event, individuated by the presence of intention as much as the former one is individuated by the absence of intention, for which Gunnar is held responsible. The reason why the question of intention is so important to the individuation of the events is precisely that it tends to be assumed (rightly or wrongly) that behaviour which is caused by my intentions is behaviour with regard to which I am free,

behaviour which I could refrain from adopting and behaviour, therefore, for which I am to be held responsible. This would not be so in the case of behaviour which is caused by Cosser's manipulating me. If free will is false, of course, then it is not true that I could act other than how I do in the case of behaviour which is caused by my intentions, but I repeat that intention is only regarded as a criterion of moral responsibility (and hence a criterion of event-individuation in the case at issue here) because of the assumption that we are free with regard to what our intentions are. If I am not, in fact, free with regard to what my intentions are, then I would think there no reason to hold me responsible for behaviour which is caused by my intentions. Despite the fact that Gunnar shoots Ridley in both cases, the event for which he is held responsible is a different event from that for which he is not so held: "shooting Ridley as the effect of the intention to shoot Ridley" is something Gunnar was (perhaps) able to refrain from doing, and is consequently responsible for it, whilst "shooting Ridley as the effect of Cosser's causing" is something Gunnar is unable to refrain from doing, and something, consequently, for which he is not responsible. This is paralleled in the Honderich example at the beginning: although in some sense of the words, the movement is the same, a knee-jerk reflex is a different event from the moving of one's lower leg. The latter case, as Honderich says, is something which is the effect of intention, and is consequently something for which I may (at any rate) be responsible, whilst the knee-jerk is not an effect of intention, is clearly a case where I could not have done otherwise, and consequently a case where I am not responsible. I think that this kind of move therefore, does nothing to counter-evidence PAP. I see no reason whatever to believe PAP false. And indeed, the comments of Searle which I cited earlier, should leave us in no doubt as to the centrality of intention in action-individuation. (Whilst I am confident of the truth of PAP, I may have seemed to suggest a more watertight connection between responsibility and intention than there probably is. For an excellent discussion of some confusions in this area, see Peter A. French's essay "Fishing the Red Herrings Out of the Sea of Moral Responsibility"[2]). I believe PAP to be true

and since I believe it likely that we have free will, I think it reasonable to suppose that we are responsible for some of our actions.

Notes

1. In Honderich, "One Determinism".

2. In Lepore & McLaughlin, Actions & Events: Perspectives on the Philosophy of Donald Davidson.

BIBLIOGRAPHY

M.R. Ayers: The Refutation of Determinism. (Methuen & Co. Ltd., 1968).

Donald Davidson: Essays on Actions and Events, (Clarendon Press, Oxford, 1980).

Daniel Dennett: Brainstorms (The Harvester Press Ltd., 1978).

Ted Honderich: Essays on Freedom of Action (Routledge & Kegan Paul Ltd., 1973).

John Hospers: Human Conduct (Harcourt, Brace & World Inc., 1961).

Keith Lehrer: Freedom and Determinism (New York, 1966).

Ernest Lepore & Brian McLaughlin: Actions & Events: Perspectives on the Philosophy of Donald Davidson (Basil Blackwell, 1985).

J.L. Mackie: Ethics: Inventing Right and Wrong (Harmondsworth, 1977).

J.L. Mackie: Persons and Values (Oxford Clarendon, 1985).

G.E. Moore: Philosophical Papers (London, 1959).

Thomas Nagel: The View From Nowhere (Oxford University Press, 1986).

Friedrich Nietzsche: The Gay Science, translated by Thomas Common (T.N. Foulis, 1910).

Gilbert Ryle: The Concept of Mind (Hutchinson, 1949).

John Searle: Intentionality (Cambridge University Press, 1983).

Wilfrid Sellars: Science, Perception and Reality (London, 1963).

Peter Van Inwagen: An Essay on Free Will (Clarendon Press, 1983).

Gary Watson (ed.): Free Will (Oxford University Press, 1982).

The Poems of Wordsworth, ed. Thomas Hutchison, M.A. (Oxford University Press, 1926).

Journals

Analysis, 1982 (Basil Blackwell, Oxford).

Mind, 1945 (MacMillan & Co. Ltd., London).

Midwest Studies in Philosophy, 1981 (University of Minnesota Press).

Philosophical Review, 1949 (W.F. Humphrey Press Inc., Geneva, N.Y.).

Philosophical Review, 1974 (ed. Faculty of the Sage School of Philosophy, Cornell University, Ithaca, New York).

Synthese, 1974 & 1987 (D. Reidel Publishing Company, Dordrecht, Holland).